

NORTH CAROLINA SCIENCE AND TECHNOLOGY RESEARCH CENTER
A UNIT OF THE DEPARTMENT OF NATURAL AND ECONOMIC RESOURCES

**A STUDY OF THE RELATIVE EFFECTIVENESS AND COSTS
OF COMPUTERIZED INFORMATION RETRIEVAL
IN THE INTERACTIVE MODE**

by

**Frederick O. Smetana
Mary Ann Furniss
T. Robert Potter**

PRICES SUBJECT TO CHANGE

NOVEMBER 1974

(NASA-CR-142206) A STUDY OF THE RELATIVE EFFECTIVENESS AND COST OF COMPUTERIZED INFORMATION RETRIEVAL IN THE INTERACTIVE MODE (North Carolina Science and Technology Research) N75-18116
Unclas
CSCL 05B G3/82 17567

Reproduced by
**NATIONAL TECHNICAL
INFORMATION SERVICE**
US Department of Commerce
Springfield, VA. 22151

STR - 505

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM THE BEST COPY FURNISHED US BY THE SPONSORING AGENCY. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE.

A STUDY OF THE RELATIVE EFFECTIVENESS AND COSTS
OF COMPUTERIZED INFORMATION RETRIEVAL
IN THE INTERACTIVE MODE

by

Frederick O. Smetana
Mary Ann Furniss
T. Robert Potter

North Carolina Science and Technology Research Center
Research Triangle Park, N.C.

ABSTRACT

Results of a number of experiments to illuminate the relative effectiveness and costs of computerized information retrieval in the interactive mode are reported. It was found that for equal time spent in preparing the search strategy, the batch and interactive modes gave approximately equal recall and relevance. The interactive mode however encourages the searcher to devote more time to the task and therefore usually yields improved output. Engineering costs as a result are higher in this mode. Estimates of associated hardware costs also indicate that operation in this mode is more expensive. Skilled RECON users like the rapid feedback and additional features offered by this mode if they are not constrained by considerations of cost.

TABLE OF CONTENTS

	<u>Page</u>
OBJECTIVES OF THE STUDY.	1
RESEARCH METHODS	1
SEARCH QUESTION.	3
RESULTS OF THE FIRST SEARCH PROBLEM.	5
LOGICAL PROCESSES EMPLOYED BY SEARCHER WWS AS DEDUCED FROM HIS SEARCH	6
SECOND SEARCH LOGIC EMPLOYED BY WWS	20
LOGICAL PROCESSES EMPLOYED BY SEARCHER MRN.	21
LOGICAL PROCESSES EMPLOYED BY SEARCHER TRP.	32
LOGICAL PROCESSES EMPLOYED BY SEARCHER SRW.	39
LOGICAL PROCESSES EMPLOYED BY SEARCHER AWL.	44
LOGICAL PROCESSES EMPLOYED BY SEARCHER PJC.	56
LOGICAL PROCESSES EMPLOYED BY SEARCHER SFS.	63
LOGICAL PROCESSES EMPLOYED BY SEARCHER CLN.	69
CORRELATION OF SEARCH RESULTS WITH SEARCHER'S BACKGROUND	79
CORRELATION OF SEARCH LOGIC WITH RECALL ACHIEVED.	80
RELATIVE EFFECTIVENESS OF INTERACTIVE AND BATCH SEARCH MODES	81
RELATIVE COSTS OF BATCH AND INTERACTIVE SEARCH MODES	84
SUBJECTIVE USER EXPERIENCE	88
APPENDIX	

OBJECTIVES OF THE STUDY

Upon acquisition of a RECON terminal, the North Carolina Science and Technology Research Center undertook a series of experiments designed to illuminate

1. The effectiveness (as measured by recall and relevance) of the interactive search mode as compared with the batch search mode.
2. The differences in costs between searches performed interactively and those performed in batch processes.
3. The search methods employed and results achieved by searchers of varying background and searching experience.

RESEARCH METHODS

The last objective was attacked first. Each searcher was put before the cathode ray tube display and typewriter keyboard. A printer alongside was rigged to record the entire "conversation," including clock time, between the searcher and RECON. The searcher was handed a statement of the problem and asked to go to work. The room was semi-sealed against disturbance. Only if the searcher had trouble making himself understood to the system (which sometimes happened with inexperienced searchers) or encountered technical difficulty (which occurs at random times) did he emerge and ask assistance of monitoring personnel.

The search problem selected was one received during the normal course of business at the North Carolina Science and Technology Research Center. It had proven to be particularly troublesome because the 400 or so relevant documents in the collection suffered from indexing vagaries that made their retrieval difficult without, at the same time, retrieving a large quantity of extraneous material. In such a problem, it was felt, a searcher could really demonstrate his skill and knowledge. The search problem statement is reproduced at the end of this section. The printed record of the "conversation" with the computer was given to the experiment director along with the "hit" list the searcher felt was his best effort. Abstracts for all of those "hits" were examined by the experiment director who decided which of the hits were relevant. This was done to determine what portion of the searcher's hit list was pertinent and how many of the pertinent documents in the entire collection the search identified.

The printed records of the "conversations" with the system were then prepared in the form of a standardized flow chart to facilitate identifying the approach and results of each searcher. In the process of preparation, mistakes which elicited responses from the system such as "Invalid Command Proceed" were eliminated as were garbled conversations resulting from system failures. Reproduced as Figure 1 (see end of this section) is the flow

chart for a search performed by an undergraduate engineering student with a fair knowledge of the scientific fundamentals involved and a better knowledge of system operation and indexing problems. The search shown in Figure 1 required the searcher to remain at the terminal approximately three hours. Figures 2-8 present flow charts depicting the search procedure used in the other seven searches made on this same question.

Several approaches were tried in order to develop data toward the first objective. The same individual was asked to make his search first by the batch process and then, before the results of his efforts were known, do it again on RECON. The outputs and strategies were then compared by the experiment director. A limited number of searches, done by one person on RECON and another by batch process, were tried but the results there seemed more dependent on individual approaches than on the differences between RECON and the NC/STRC batch search procedure. This approach was therefore discontinued.

Costs differences between a batch process search and an interactive search were also very difficult to evaluate for a number of reasons:

1. The line charges and computer charges per search on RECON are not directly known because of the way the system and its accounting are set up;
2. An individual usually does not perform a search exactly the same way on the two systems.
3. RECON experiences occasional shutdowns and frequent over-loadings of the system which result in long response times to questions and instructions.
4. An engineer can and usually does obtain intermediate output during a RECON search which he uses to phrase his remaining questions.

Although discussing the cost differences between the batch and interactive search approaches was judged to be a difficult task for the reasons mentioned above, it was felt that STRC had enough specific accounting data on the batch system and could obtain enough data on the other RECON-type on-line search systems to make a worthwhile cost comparison between the two modes of searching. In addition, the experimental RECON searches provided sufficient data on the amount of human effort involved in the interactive search mode and STRC cost accounting reports provided the same information for batch searches. The resultant cost comparison can be found in another section of this report entitled "Relative Costs of Batch and Interactive Search Modes."

In addition to these tests, members of the technical staff of NC/STRC used RECON as they wished during normal Center operations. As their skills and knowledge of the system improved, various members began to make increasing use of it. Included herein is a report of one subjective reaction to RECON - what he liked, what he did not like, and how he liked to use it.

SEARCH QUESTION

"Airborne Thermal Mapping of Water Surfaces"

A consulting firm is bidding on a contract to map the water temperature in lakes which are used to cool thermal power plants. The mapping will be done by flying over in an airplane (3000' - 6000' above lake surface), observing the lake with a radiometer, recording the radiometer output, and then processing the data so as to be able to draw lines of constant temperature on an accurate outline of the lake. The purpose, of course, is to make it possible to check compliance with thermal pollution standards easily and quickly.

Specifications call for measuring accuracies of 0.5°F. The consulting firm thinks that 1°F is about all that they can do realistically. So, they would like to know (a) does any responsible group claim 0.5°F. accuracy? (b) If so, what

1. sensor technology
2. mathematical modeling techniques
3. data reduction methods
4. computerized data processing techniques

do they employ?

NOTE: A radiometer is a device which senses the heat radiated from the surface of a body. It can be focused so that it "looks" at a relatively small area. The atmosphere absorbs some of this radiation, the absorption being dependent upon air temperature, density, and distance the radiation travels. Therefore, the angle at which the radiometer looks at the surface from the airplane determines how much air the heat waves have to go through, what its density is, and how it's "layered."

From this knowledge and a mathematical model of the absorption and dispersion of infrared radiation by a variable density atmosphere, one can arrive at an estimate of the surface temperature in terms of the radiometer indication. Also, a lake is only a semi-opaque body, which means that all of the radiation that one detects is from the layer immediately below the surface. These effects are also included in the mathematical model.

The radiometer output is an electrical signal proportional to the intensity of the thermal radiation and its spectral distribution (how much energy is radiated at each wavelength). One must therefore have a procedure for converting this signal to temperature. Ideally, the whole process could be one in which the user merely submits a magnetic tape recording of radiometer output looking at a given part of the lake to a computer program. The program reduces the data from this and other passes over the lake, and from these maps the lake surface temperatures.

RESULTS OF THE FIRST SEARCH PROBLEM

The search question posed was first performed by FOS as a batch search. Several reruns and considerable expense were required to obtain a sufficient quantity of relevant output. Center management felt that it would therefore be a good question on which to compare the effectiveness of various approaches to the question and the backgrounds of the searchers. The capability of recording the dialog which the searcher had with the system made it possible to study in detail the logic used by each individual. From this dialog, more formalized flow charts were prepared. The seven flow charts are discussed in detail below. Following this is a correlation of the relevancy and recall of the various searches with search logic and searcher's background. Completing the data assembled from this experiment is Table 1 (see Appendix). This lists the documents recovered from all searches which the experiment director called pertinent, the searches identifying that document, the number of pertinent documents identified by each search as a percentage of the total pertinent documents (recall) and as a percentage of the hits found by that search (relevance). Also shown in Table 1 are 10 additional pertinent documents found by the experiment director by constructing a search from the index terms appearing most often in the highly pertinent documents. This post experiment search had 103 hits of which 45 were pertinent. Thirty-five of these had been recovered during the experiment.

In passing, it may be noted that manual review of each search output in sequence will usually lead to some inconsistencies in identifying documents which had been identified as pertinent in one search, but were not so identified by the experiment director in his review of the search indicated. Between 6% and 30% of the pertinent hits in each search were so affected. In a topic of this nature many documents will be of at least peripheral interest to a broadly-trained, innovative researcher. However, as one moves further and further from the central thrust of the subject, discrimination becomes fuzzy. Whereas today a particular document with a tenuous relation to the subject might seem of possible interest, tomorrow, it may not. As a result of this ambiguity, in subsequent comparisons of searcher effectiveness a master hit list was compiled by keypunching the output of each search - along with an identifying code to indicate the source search - and then sorting. Only the abstracts from the master hit list were then examined to select the relevant documents. The relevant document accession numbers were then keypunched and inserted in a computer program which gives the recall and relevance for each individual search.

LOGICAL PROCESSES EMPLOYED BY SEARCHER WWS AS
DEDUCED FROM HIS SEARCH

In order to follow the logical processes depicted in Figure 1, one should note that the following symbolism is employed:

(1) A rectangle indicates the term which was the basis for a following operation.

(2) In the upper left corner of the rectangle is a symbol denoting the source of the term:

m	from his mind
e	from a previous expansion
t	from a thesaurus expansion
c	from a document's bibliographic citation

(3) A rectangle with curved sides attached indicates what the searcher saw on the CRT screen, e.g., a "display."

(4) The elongated hexagon denotes an operation, e.g., expand "thermal."

(5) The diamond indicates a decision which can be answered yes or no.

(6) The circle indicates premature termination of an operation.

(7) Operations, display, etc., are arranged in chronological order with arrows leading from the proper antecedents and to the point in the search where subsequent operations are performed.

(8) Where it is necessary to break a line at the bottom of a column an identifying letter is inserted so that when the reader again sees that letter he knows the line of which it is a continuation.

(9) Lines to the right of terms in a display indicate that the searcher selected those terms to form sets.

The searcher began by requesting an expansion about the term he considered to be basic to the whole search, the term "thermal." On the basis of the first display⁺ he selected two terms and felt that these were sufficiently pertinent for him to continue the displays until the display limit⁺⁺ was reached. From this expansion he selected a total of 9 terms

⁺ Display refers to the material presented at a given time on the face of the cathode ray tube viewing device.

⁺⁺ Approximately 14 additional terms are displayed on the CRT each time the command "more" is given until the limit of 41 terms has been reached.

and formed 8 sets. Note that he is assured of a large number of documents upon which he could perform further logical operations by the fact that the term "thermal" is posted to 10,954 documents.

One of the terms displayed in the expansion, "thermal emission," seemed, on the basis of the problem statement and his general knowledge of the subject area, to be worth examining further. The searcher therefore requested a thesaurus expansion from which he selected three terms. He then combined these to form set 9.

To get another start on the problem, the searcher went back to the central idea and selected a broad but related term, "temperature," on which to request an expansion. He knew from previous experience that asking the same question in several ways is often necessary to assure that documents dealing with a particular subject (but indexed by different individuals using similar but not identical terms) are retrieved. Unfortunately, he misspelled it "temperture." The display indicated no postings under this term. The searcher nevertheless did find what he considered to be a useful term in this display. He then decided to overcome his spelling deficiency by using first the root of the term, "tempera." This produced the desired result and he again displayed all the terms up to the limit, selecting 12 to form 10 additional sets.

Next, he chose to perform a thesaurus expansion on one of the terms which appeared during the previous display, "temperature measurement." This term, of course, is specifically what the search is about. Since the measurement was to be carried out by observing the radiation, he chose "radiation pyrometer" as a set.

Since the search statement had contained the term "radiometer," he then decided to see what useful terms might be alphabetically related to this. The expansion was reasonably productive but he decided on viewing the first display that a thesaurus expansion would be more appropriate. From this he selected seven terms.

He then went back to the general concept of "measurement" and performed a thesaurus expansion which yielded a number of useful terms.

In the general sense the concepts of "temperature" and "heat" are closely related. Seeking to formulate the idea of "temperature measurement by detection of radiant heat flux" in other terms, he chose to expand the general term "heat." Only three of the terms displayed, he felt, added significantly to those he had already selected.

He then approached the idea of measuring the temperature distribution over a surface as a form of mapping and sought to determine whether any of the related terms would be useful. From the thesaurus expansion he selected only three.

He then tried a different approach on the idea of measurement by using the term "sensing." This appeared to be quite productive. At this point he decided to winnow down the number of documents to those more pertinent. He combined those sets dealing with temperature measurement

to get a single set, number 39. He also combined sets dealing with radiometers to obtain set 40 and sets dealing with mapping and sensing to obtain set 41. The sets dealing with temperature in a general way were intersected with set 41 to give a new set, 42, which contained 465 documents indexed under both the idea of temperature and that of mapping. The general idea of measurement was then combined with set 41. The new set, number 43, was intersected with sets on heat and radiation measurement to obtain 1443 documents dealing specifically with radiation mapping. Set 43 was also intersected with sets 39 and 40. The result was a set of 1473 documents dealing with mapping by radiometer.

In this process, one term with only two postings caught his eye: "sensor for airborne terrain analysis." He decided therefore to display the two citations to see how relevant they were. They were not.

Up to this point he had established that the documents in sets 39, 40, 42, 44, and 45 with some 9000 postings seemed to bear on the problem of measuring or mapping radiation fluxes over large areas. Since he was specifically interested in the emission from water he developed some expansions about the idea of liquids. Except for the general terms themselves, the results did not appear to him to be particularly promising. He then began to explore the idea of surfaces and assembled all the documents (495) dealing with the idea of liquid surfaces.

Backtracking for a moment, he intersected the idea of radiometers (set 40) with the concepts of temperature measurement, temperature mapping, and radiation mapping (sets 39, 42, and 44). This produced set 53 with 347 accessions. Ultimately, he decided that set 53 represented his best effort on the question, but at the time, he decided to try one more tack to see if some useful documents might show up.

He had not yet investigated the possibilities of approaching the question from the point of view of sensing from an aerial platform, which is how this job was to be done in the field. He began to travel this path by asking for an expansion on airborne. Although he selected one term, the expansion did not appear to be too fruitful. He then made a thesaurus expansion on the selected term without gaining what he considered to be additional insight. At this point he tried an expansion on a general, but related, term, "aerial." Again he got only one term which he felt might be useful, so that he decided not to continue the expansion.

The two terms selected from this effort he combined to create set 56. He was obviously undecided as to what to do with it at the moment because he then intersected set 53 with set 52 (dealing with liquid surfaces) and got only 4 documents (set 57). This of course is too few, so he then tried searching the 2956 accessions in set 56 for index terms also in set 53 (by performing a logical intersection or "and"). This gave set 58 with 31 documents.

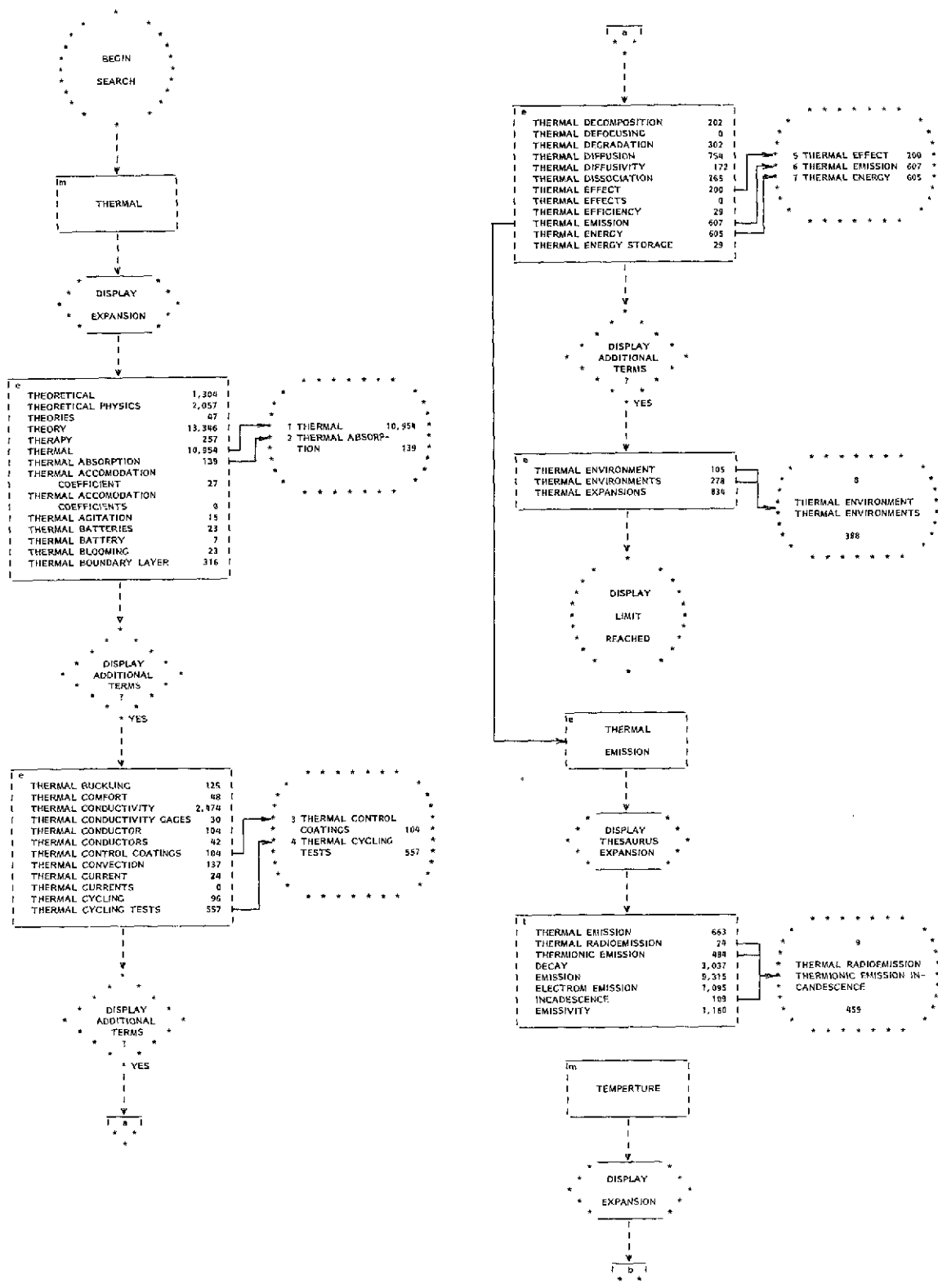
To see if he were in the right area in his selections, he asked for two of the citations from set 57. These looked to be very pertinent, so he intersected set 57 with set 58. Since three of the four documents in

set 57 also appeared in set 58, he felt that set 58 was probably also quite pertinent. He therefore asked that complete bibliographic information on the documents in 57 and 58 be printed. Since both were subsets of 53, he felt that 53 would probably also contain a large amount of relevant data. Finally, then, he asked that all the accession numbers in set 53 be printed.

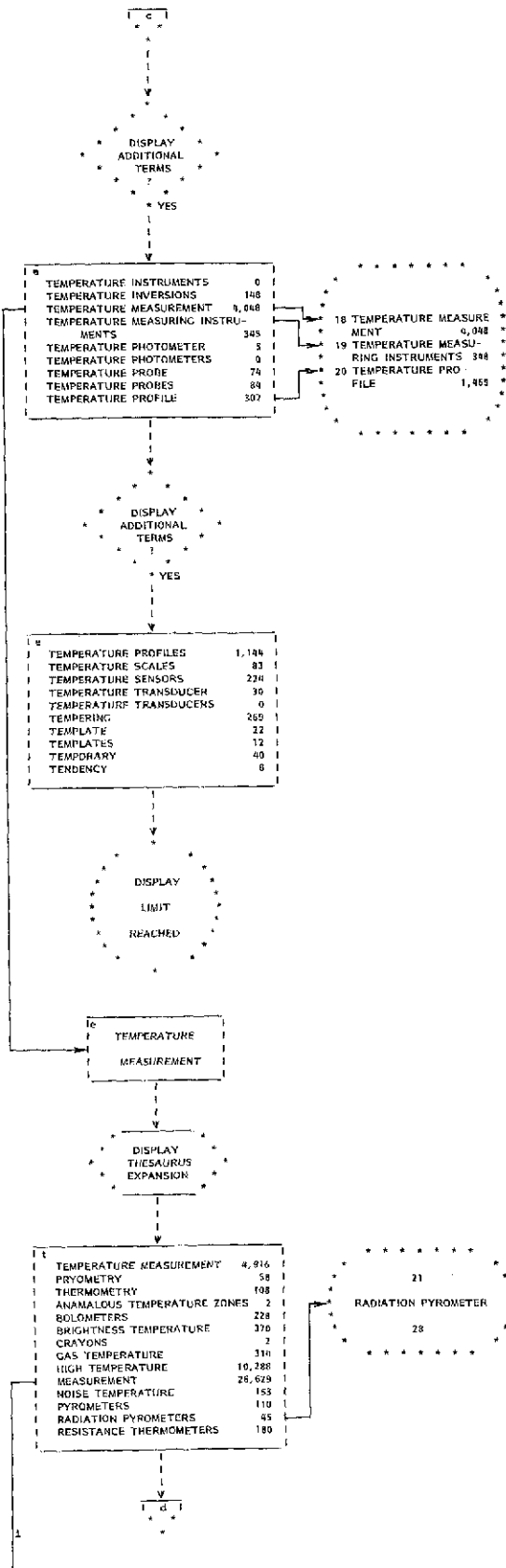
As a matter of interest it may be noted that ultimately 388 separate relevant documents were found in the system through a total of eight searches by different individuals. The particular search discussed here found 121 or 31.2% of these. Of the 347 hits designated by the searcher as constituting his search results, 34.9% were relevant. This is very good performance for this type of question (i.e., one without highly specific indexing).

Of particular note also is the fact that this search found 55 relevant documents turned up by no other search. The additional 66 relevant documents were found in one or more of the other searches. Only 15 of the 388 relevant documents appeared in 5 or more searches. Only 5 relevant documents appeared in 6 or more of the searches and of these, 3 appeared in 7 searches. No document appeared in more than 7 searches.

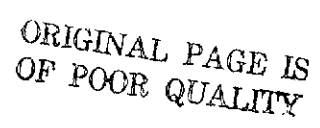
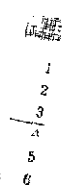
FIGURE 1



ORIGINAL PAGE IS
OF POOR QUALITY



ORIGINAL PAGE IS
OF POOR QUALITY.



CONSISTENCY	87
COUNTING	379
DATA	17,041
DEFINITION	979
DENSITY MEASUREMENT	1,229
DEPTH MEASUREMENT	209
DETECTION	5,826
DILATOMETRY	173
DIMENSIONAL MEASUREMENT	222
DOWNRANGE ANTIMISSILE MEASUREMENT PROGRAM	60
DOWN RANGE MEASUREMENT	29
DRAC MEASUREMENT	273
ELECTRICAL MEASUREMENT	1,570

DISPLAY
ADDITIONAL
TERMS
YES

ELECTROMAGNETIC MEASUREMENT	715
ELECTROMAGNETIC NOISE MEASUREMENT	22
ESTIMATING	774
EVALUATION	5,700
EXAMINATION	337
FLOW MEASUREMENT	1,594
FREQUENCY MEASUREMENT	834
FRICTION MEASUREMENT	389
GEOMETRY	5,367
GRAVIMETRY	755
HEAT MEASUREMENT	777
HIGH ALT TARGET AND BACKGROUND MEASUREMENT	11

DISPLAY
ADDITIONAL
TERMS
YES

LIQUIDITY MEASUREMENT	260
IDENTIFYING	570
INTERNATIONAL SYSTEM OF UNITS	25
LATITUDE MEASUREMENT	168
LONGITUDE MEASUREMENT	50
MACROSCOPIC EQUATIONS	195
MAGNETIC MEASUREMENT	1,127
MEASURING INSTRUMENTS	3,304
MECHANICAL MEASUREMENTS	201
METROLOGY	380

DISPLAY
ADDITIONAL
TERMS
YES

9

MONITORS	1,278
OPTICAL MEASUREMENT	2,029
PNEUMOGRAPHY	68
PRESSURE MEASUREMENTS	1,923
PROVING	245
PUPILLOMETRY	33
RADAR MEASUREMENT	937
RADIATION MEASUREMENT	2,535
RADIOACTIVE AGE DETERMINATION	381
RANGE FINDING	516
SIGNAL MEASUREMENT	876
OPTOMETRY	82
PHOTOGRAPHIC MEASUREMENT	231

DISPLAY
ADDITIONAL
TERMS
YES

SIZE DETERMINATION	394
SOUNDING	2,022
SPHYMOGRAPHY	26
STANDARDS	1,298
SYNOPTIC MEASUREMENT	229
TEMPERATURE MEASUREMENT	4,916
THRUST MEASUREMENT	344
TIME MEASUREMENT	1,114
TRAJECTORY MEASUREMENT	206
UNITS OF MEASUREMENT	148
VELOCITY MEASUREMENT	2,195

DISPLAY
ADDITIONAL
TERMS
YES

VIBRATION MEASUREMENT	866
WEIGHT MEASUREMENT	132
WIND MEASUREMENT	1,094

HEAT

DISPLAY
EXPANSION

HEART MINUTE VOLUME	57
HEART RATE	816
HEART VALVES	13
HEARTH	11
HEARTHS	3
HEAT	10,379
HEAT ACCLIMATIZATION	8
HEAT BALANCE	466
HEAT BUDGET	61
HEAT CAPACITY	152
HEAT CONDUCTION	51
HEAT CONTENT	31
HEAT DISSIPATION	66
HEAT DISSIPATION CHILLING	0
HEAT EFFECT	51
HEAT EFFECTS	0

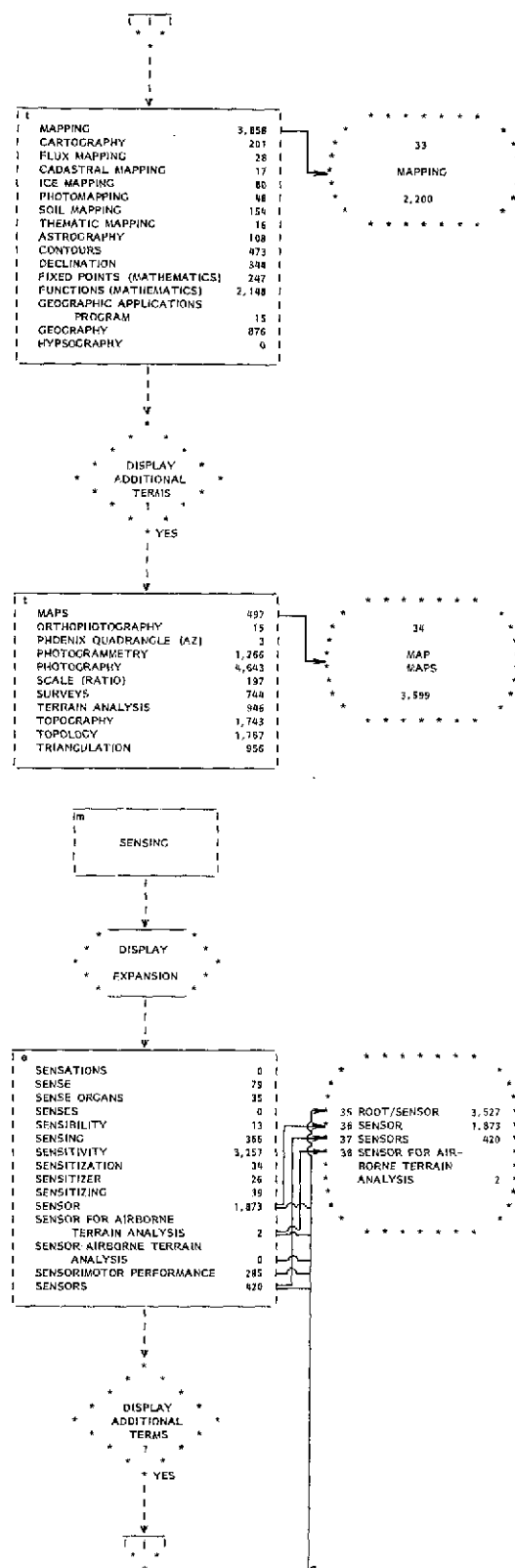
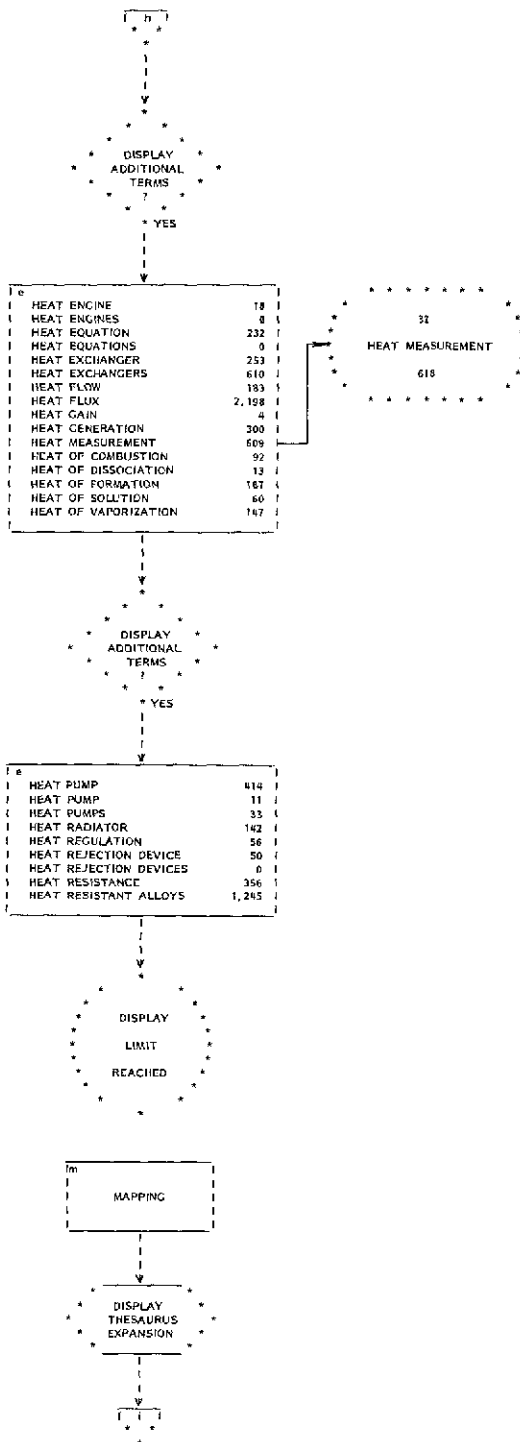
9

29
RADIATION MEASUREMENT
2,180

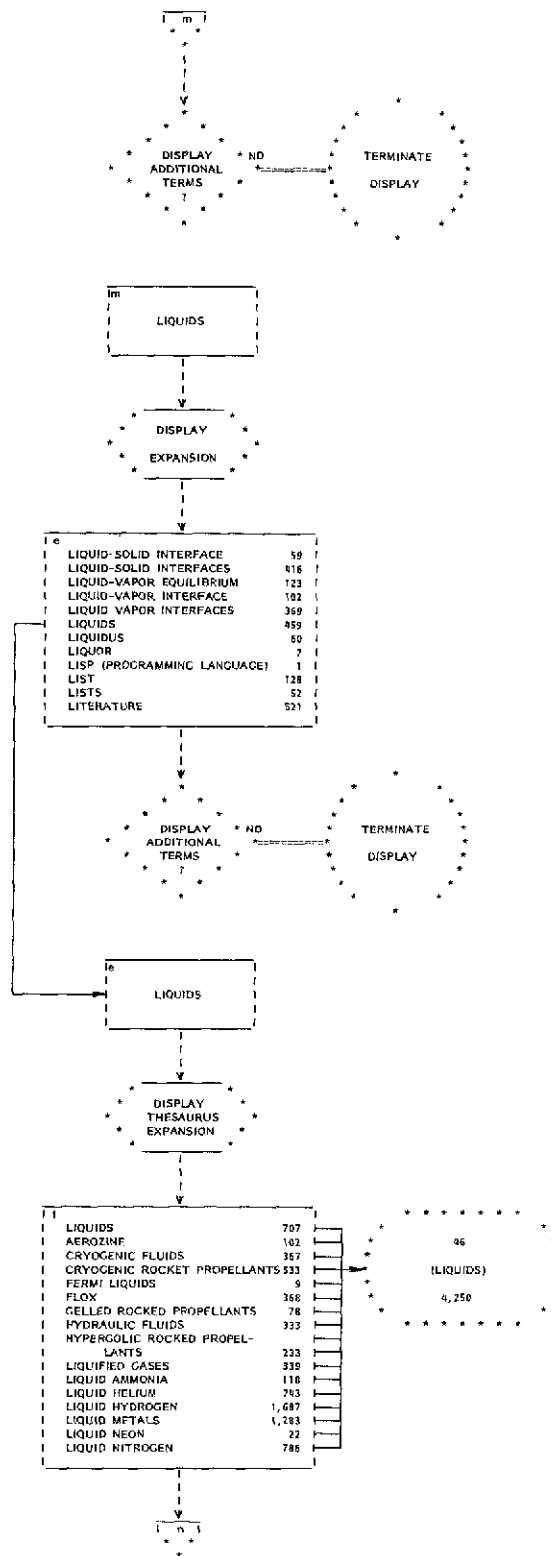
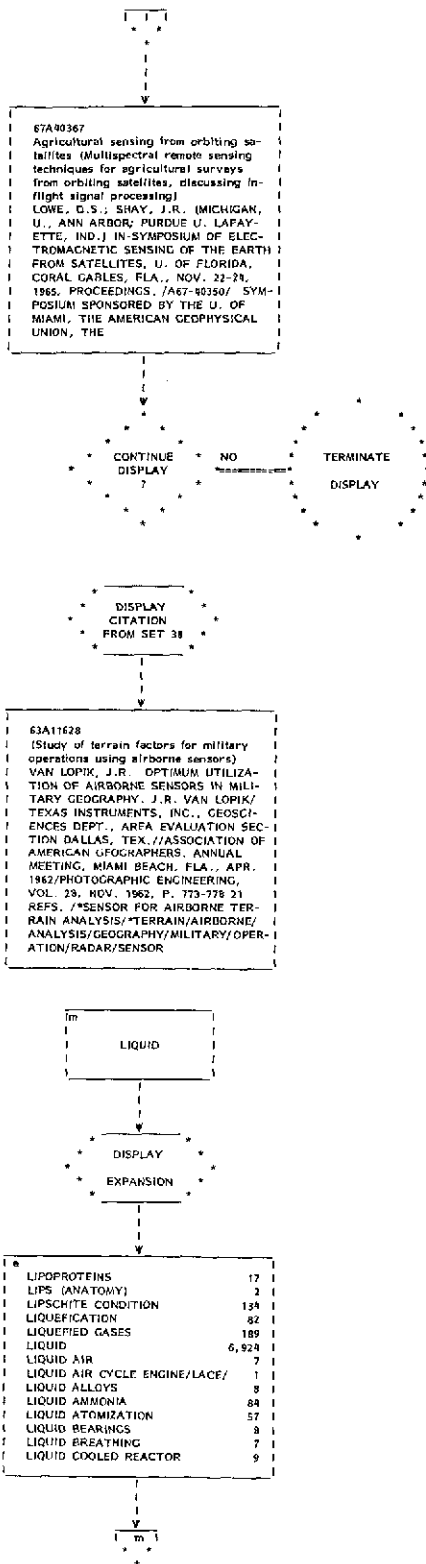
27 HEAT MEASUREMENT
28 HIGH ALT TARGET AND BACKGROUND MEASUREMENT
618
2

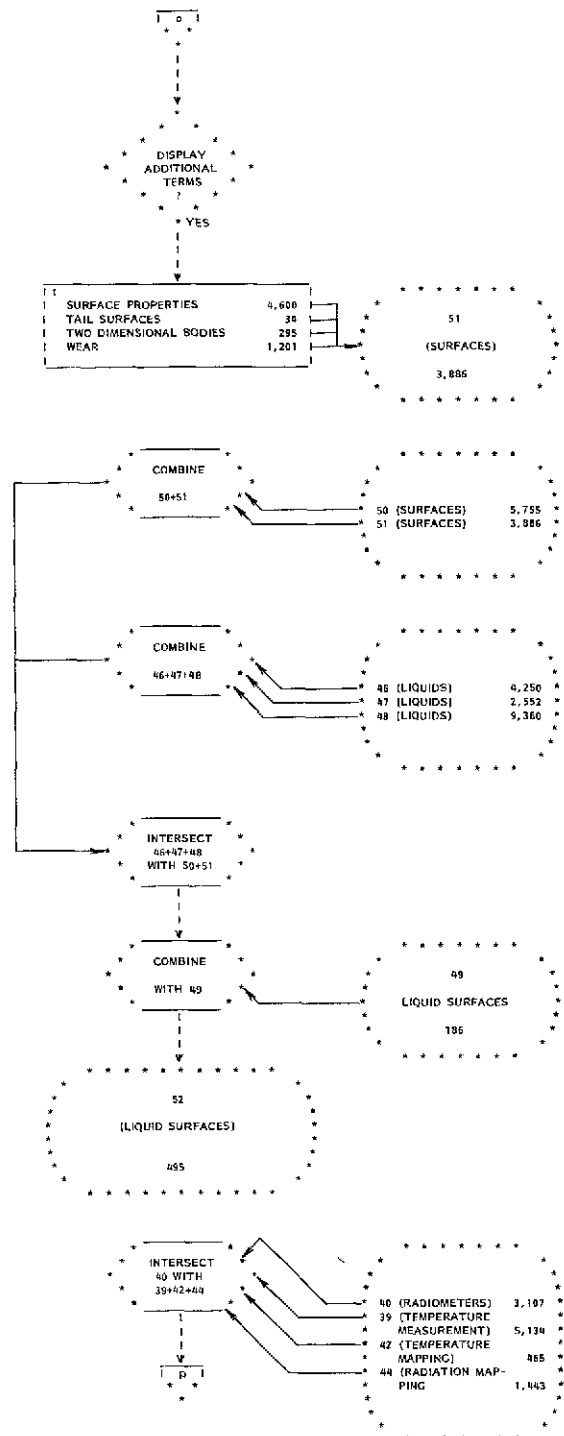
30 HEAT
31 HEAT CONTENT
10,375
31

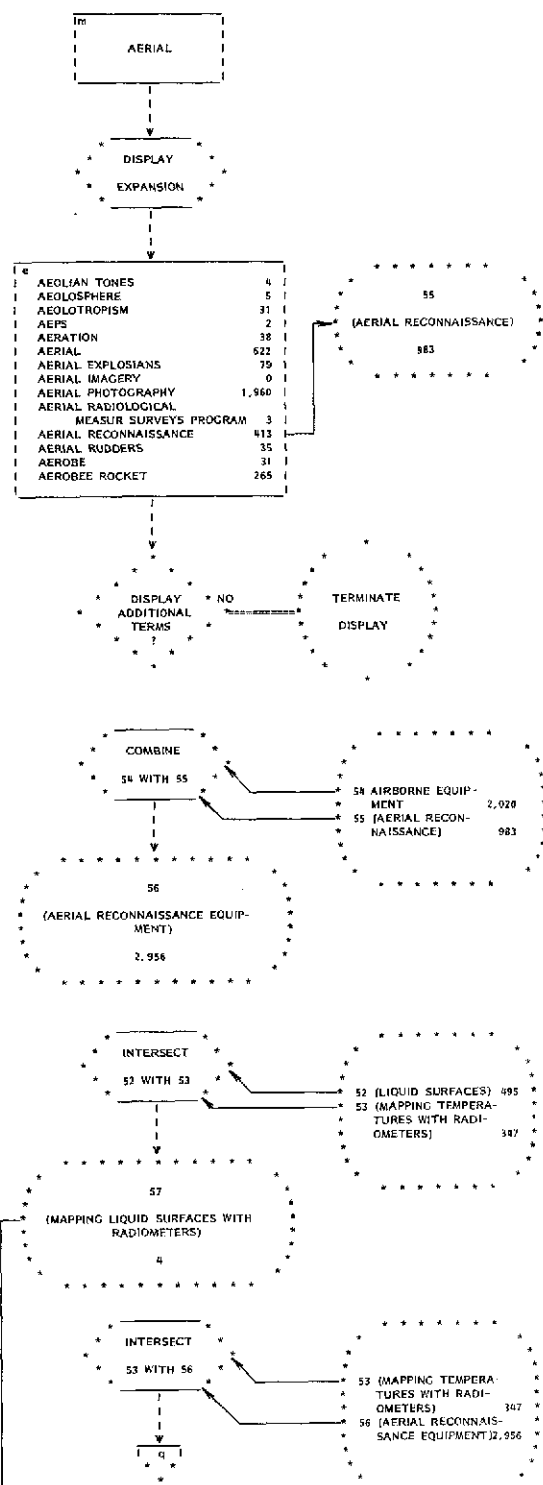
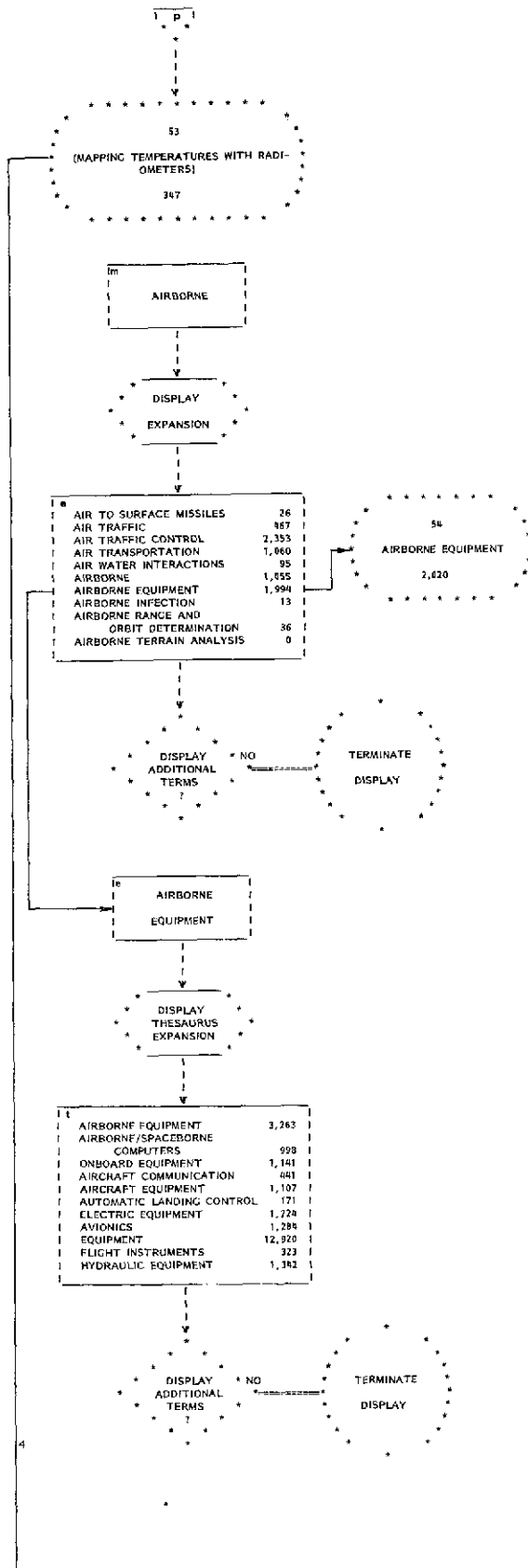
ORIGINAL PAGE IS
OF POOR QUALITY

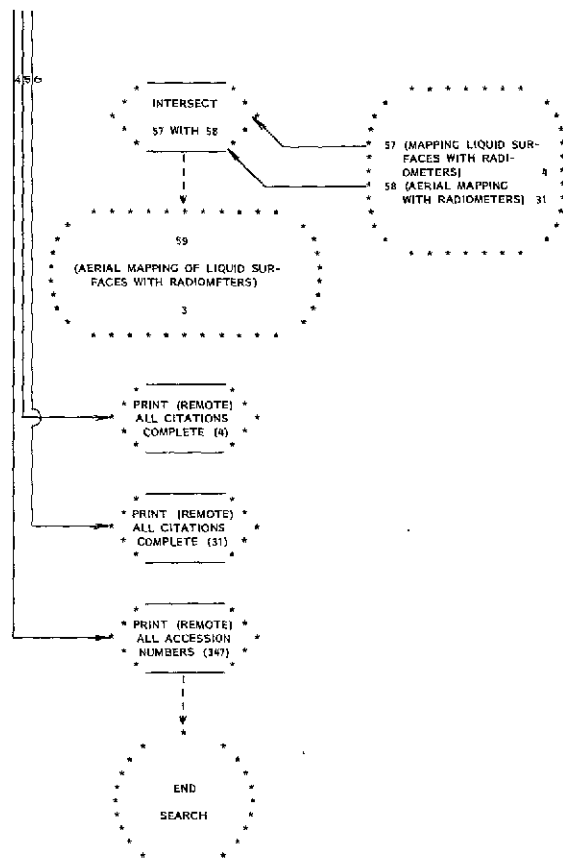
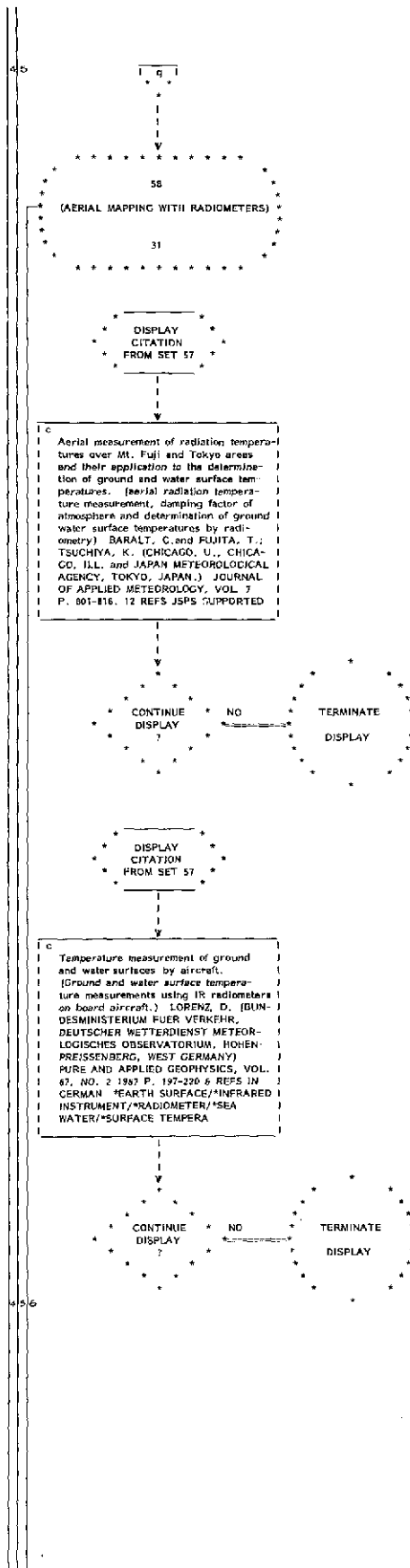


ORIGINAL PAGE IS
OF POOR QUALITY









SECOND SEARCH LOGIC EMPLOYED BY WWS

After completing his assigned search, searcher WWS felt that the task could probably have been done more simply by employing a single intersection between two groups of very general terms. Accordingly, he began by selecting "thermal" (12,557 postings), "temperature" (25,518 postings), "temperature gradient" and "temperature gradients" (2,428 postings), "maps," "map," and "mapping" (4,226 postings). He then intersected the last group with the union of the first three to obtain 205 hits. Interestingly enough, this search produced only 8 good hits, half of which were not found in any other search.

LOGICAL PROCESSES EMPLOYED BY SEARCHER MRN

The searcher is a trained literature chemist with specialization in the medical and biological areas. He had relatively low familiarity with RECON and its capabilities at the time the experiment was conducted. In addition, his thought processes were somewhat disturbed by several short but annoying systems failures during the course of his search.

He began (Figure 2) by asking for an expansion on the name of the instrument cited in the problem statement as being the one used to measure temperature. From this he selected three terms which he combined to form set 1. It also became obvious from examining this display that a simple alphabetical listing of the terms in the neighborhood of "radiometer" would not give him a clue as to terms functionally related to radiometers. He therefore asked for a thesaurus expansion on "radiometers." This he deemed more productive because he combined nine of the 20 related terms displayed to create set 2.

He next tried the idea of pollution, reasoning that because the radiometer was used to investigate thermal pollution, he might find a connection. The dictionary expansion seemed fruitless, so he tried a thesaurus expansion. Even though he looked at all the terms in this expansion, he decided that none was really suitable. He apparently did not regard the term "thermal pollution" as being sufficiently descriptive of measurement problems in thermal pollution to choose it.

Following this tack, he examined an expansion about the term "airplane," the vehicle on which the sensor is mounted during measurement. Since there are no postings under this term, it was obvious that a different term is used by the indexers to describe flying machines. He therefore tried "aircraft" and found some 10,000 postings. In an attempt to be a little more specific, he requested a thesaurus expansion on the term. The plethora of types displayed caused him to settle for the general term as set 3.

He then tried the idea of "sensor," a concept taken from the problem statement. From the expansion he selected only the general term for set 4. He also tried a thesaurus expansion on "sensors," selecting the general term as set 5 and "remote sensors" as set 6. To clean things up a bit he combined sets 4, 5, and 6 in set 7.

Water, being the medium whose temperature is to be monitored, then seemed worth checking out. A dictionary expansion was performed with the result that three of the terms - "water," "water temperature," and "water pollution" - were selected as sets 8, 9, and 10. On the hunch that something very specific might turn up on the intersection of "water pollution" and "remote sensors," the operation was performed (set 11) and some of the citations examined. At that time, the citations did not appear to be particularly relevant so the searcher tried another but similar tack combining sets 2, 4, and 5 dealing with sensors and radiometers to form set 12 and intersecting this with "water pollution" (set 10). The resulting set (13) had 36 hits, but an examination of one citation suggested that the proper connection had not yet been made. The citation did suggest another term which seemed to be related to the problem, "aerial reconnaissance" (set 14).

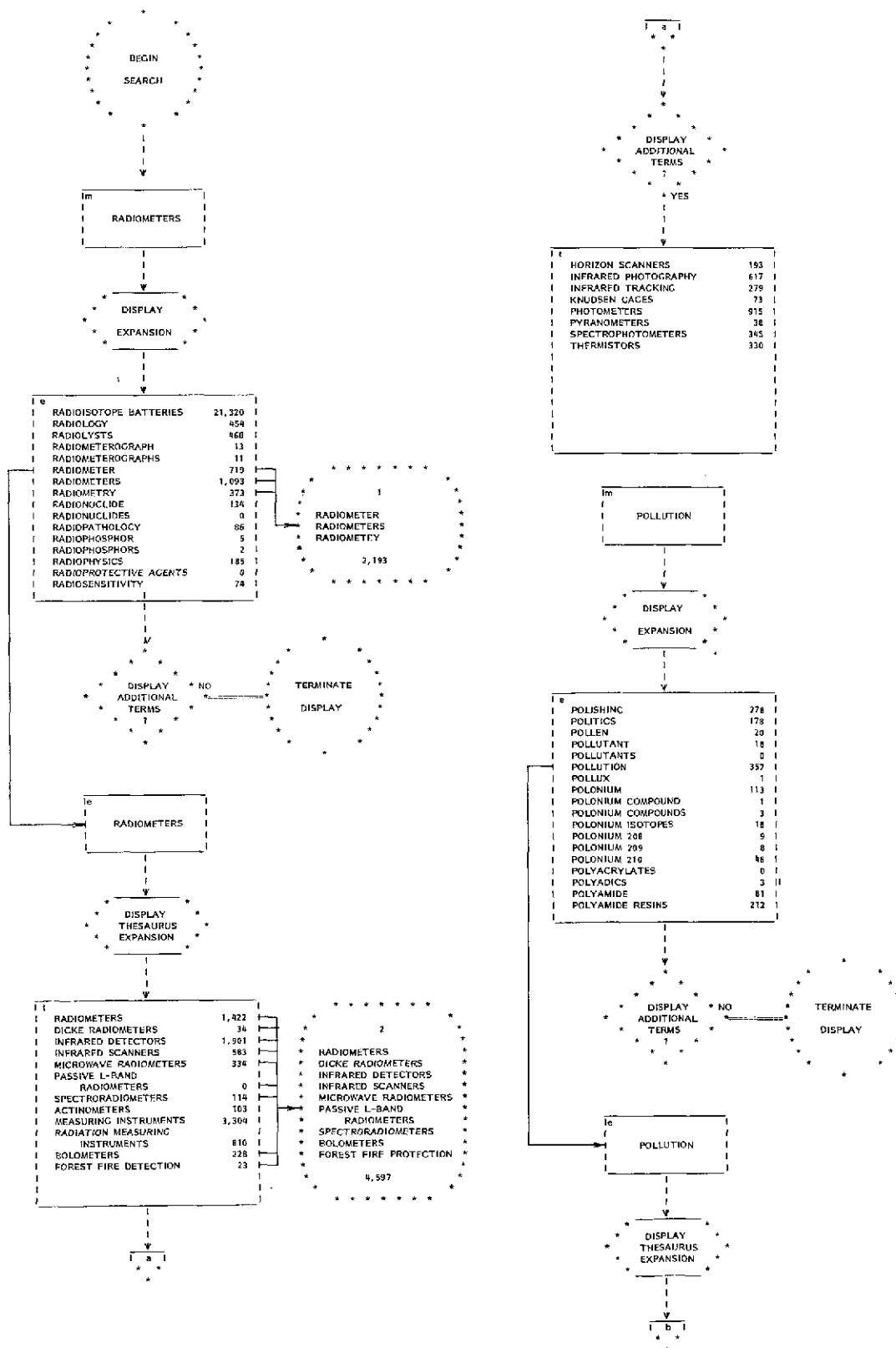
An index term from set 11 - "thermal pollution" - was then used as the basis for an expansion without yielding anything of great interest except the term itself (set 15).

The searcher then combined set 6 and 14 to collect documents dealing with aerial reconnaissance and remote sensing and searched these for those documents also indexed under "water temperature," "water pollution," and "thermal pollution." The resulting set (16) contained 126 documents. Apparently he had some second thoughts about the utility of such an intersection because the results were left to lie while he attempted a different approach: combining radiometers and sensors (sets 2, 4, and 5) and intersecting these with "water" to create set 17 with 86 hits. This he intersected with set 3, "aircraft." Only one document was recovered in this set (18). Since the number of hits was insufficient, he tried the idea of intersecting the water and thermal pollution terms with "aircraft" to create set 19 with 142 hits.

At this point he decided to look to some of the citations in sets 17, 18, and 19. Apparently set 19 appeared to stray from the subject. The others, however, seemed reasonably pertinent. He therefore combined these with 16 to create set 20 with 211 hits.

His next step was to combine sets 6 and 14 dealing with remote sensors and aerial reconnaissance and intersect the result with sets 9, 10, and 15 which had been used earlier intersected with aircraft to obtain low relevancy set 19. He seemed to hope that by employing more specific terms than aircraft he might obtain more pertinent citations. The new set, 21, contained 126 items. Apparently forgetting his history in creating set 17, he repeated the procedure to create set 22. He also repeated the creation of set 18 to form set 23. Sets 21, 22, and 23 were combined to form set 24 of 211 items and the results declared the final hit list. This list is substantially the same as set 20. 39 documents in the final hit list were regarded by the experiment director as relevant. 9 of these were not found by any other searcher. Recall was 10% of the total relevant documents in the file and 18.5% of the final hit list was relevant.

FIGURE 2



b	
↓	
POLLUTION	428
AIR POLLUTION	2,368
ENVIRONMENTAL	
POLLUTION	331
NOISE POLLUTION	159
THERMAL POLLUTION	49
WATER POLLUTION	423
CONTAMINANTS	509
CONTAMINATION	2,375
DEBRIS	474
DECONTAMINATION	613
DISSIPATION	1,592
ELIMINATION	483
ENVIRONMENTAL EFFECTS	115
ENVIRONMENTAL PROTECTION	81
ENVIRONMENTAL QUALITY	88

↓

DISPLAY
ADDITIONAL
TERMS

↓

YES

ENVIRONMENTAL SURVEYS	51
HUMAN WASTES	191
METABOLIC WASTES	111
MICROCYSTIS	2
MICROORGANISMS	745
OIL SLICKS	18
POISONS	23
PREVENTION	871
PURITY	840
QUALITY	2,734
RADIOACTIVE WASTES	329
SMOKE ABATEMENT	36
TOXICOLOGY	376
WASTE DISPOSAL	565
WASTES	59
WATER	10,485
WATER RECLAMATION	

↓

DISPLAY
ADDITIONAL
TERMS

↓

YES

WATER TREATMENT	191
-----------------	-----

↓

AIRPLANE

↓

DISPLAY
EXPANSION

↓

c

c	
↓	
AIRLINER	11
AIRLINERS	0
AIRLOCK MODULES	14
AIRMAIL	5
AIRMAT	8
AIRPLANE	0
AIRPLANE PRODUCTION COSTS	85
AIRPORT	531
AIRPORT BEACON	4
AIRPORT BEACONS	21
AIRPORT LIGHT	30
AIRPORT LIGHTS	36
AIRPORT MEDICAL SERVICE	2
AIRPORT MOBILE LOUNGE	2
AIRPORT MOBILE LOUNGES	0

↓

DISPLAY
ADDITIONAL
TERMS

↓

YES

AIRPORT PLANNING	913
AIRPORT SURFACE	
DETECTION EQUIPMENT	21
AIRPORT SURFACE	
DETECTION EQUIPMENT	
/ASDE/	4
AIRPORT TOWERS	52
AIRPORTS	803
AIRSHIPS	33
AIRSPACE	158
AIRSPEED	184
AIRWORTHINESS	40
AIRWORTHINESS REQUIREMENT	25
AIRWORTHINESS REQUIREMENTS	0
AIRY FUNCTION	115
AIRY PHASE	17

↓

DISPLAY
ADDITIONAL
TERMS

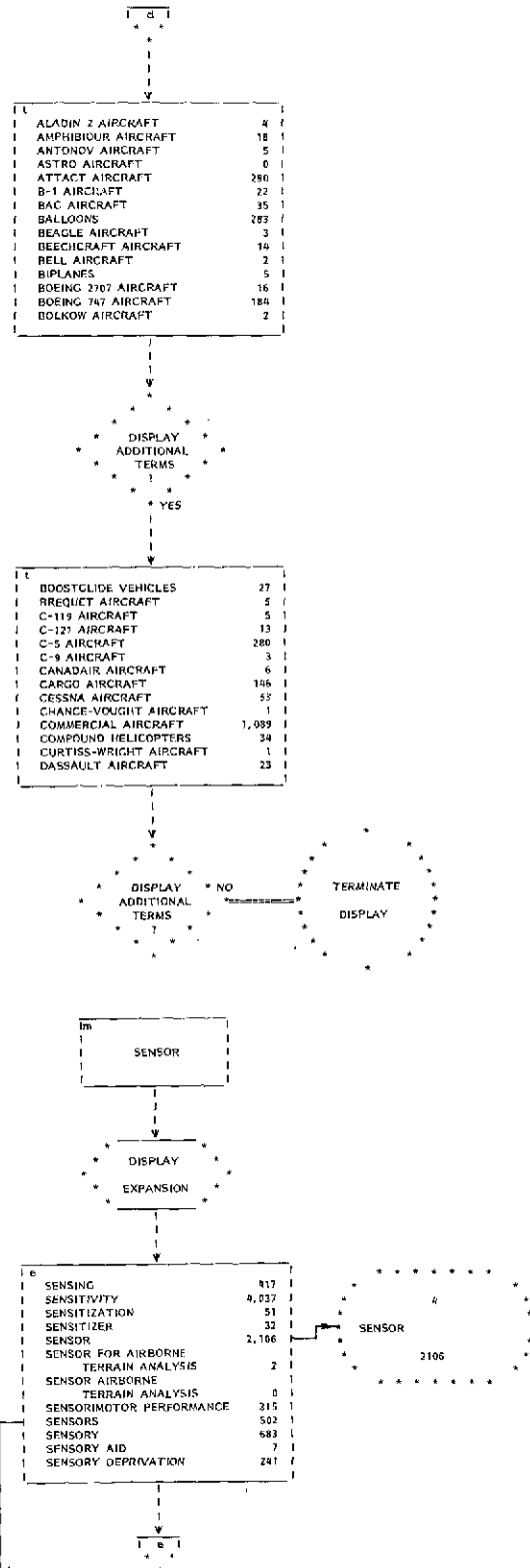
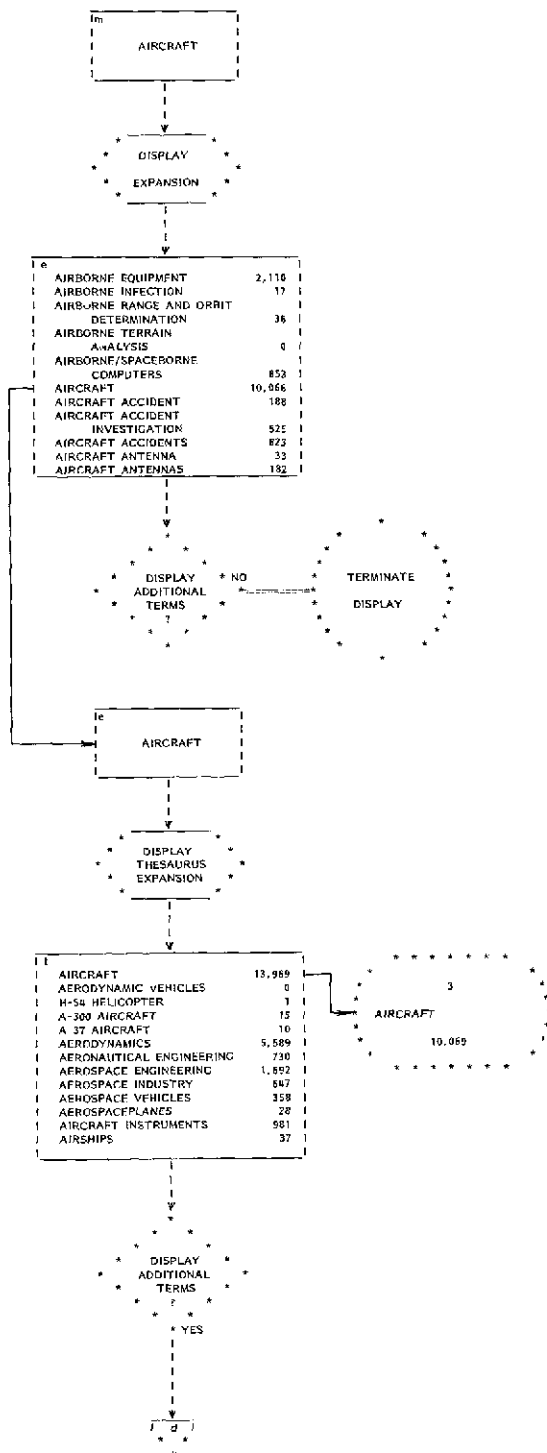
↓

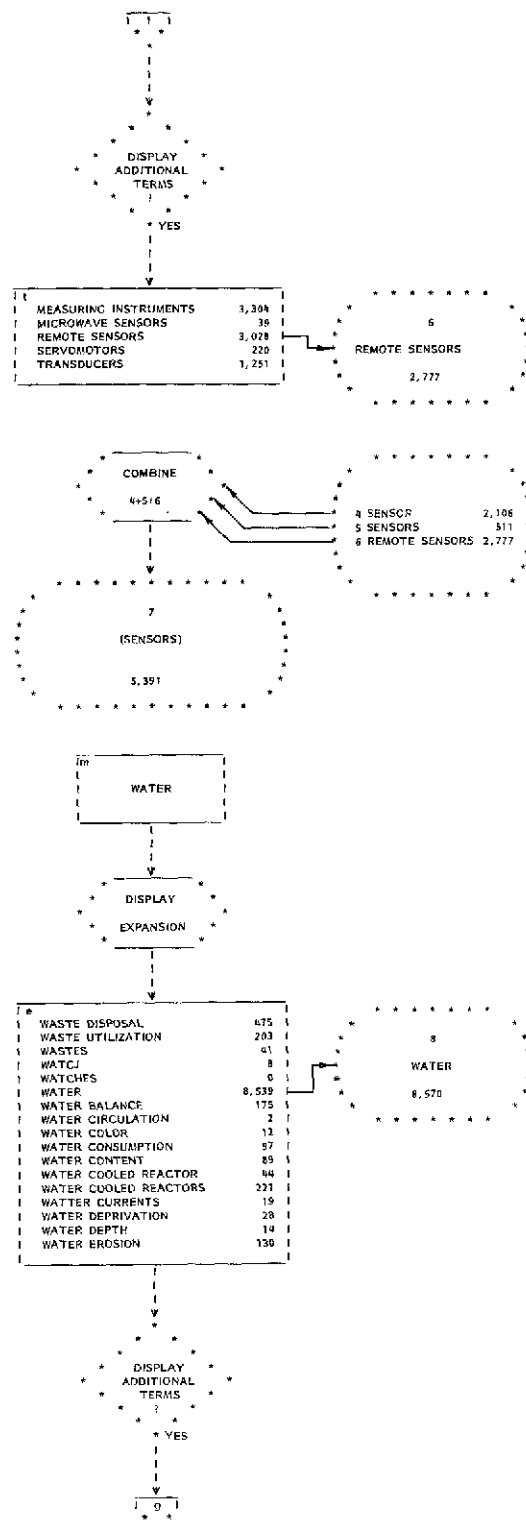
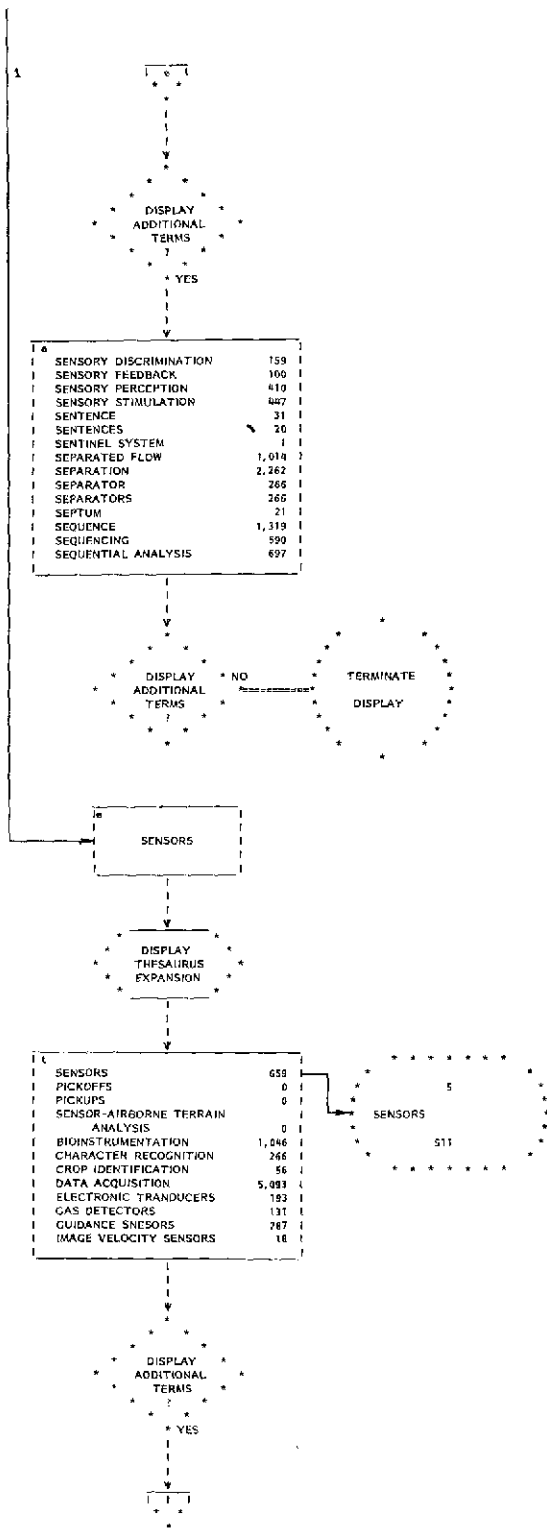
YES

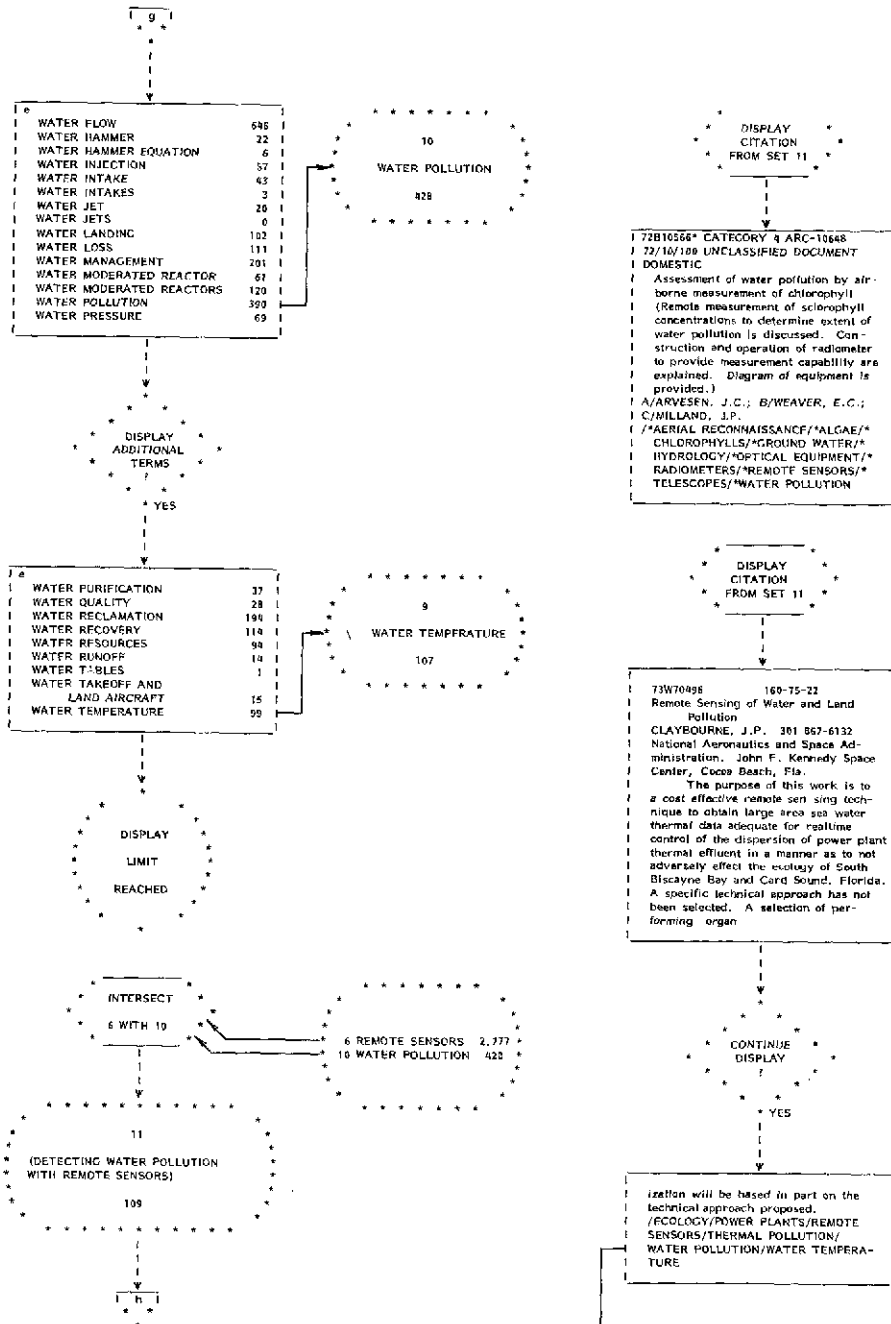
AIRYS STRESS FUNCTION	28
AJ-10 ENGINE	3
AJ-1000 ENGINE	0
AKERMANITE	3
ALABAMA	26
ALAD	1
ALADIN 2 AIRCRAFT	3
ALAS METEORITE	6
ALANF	4
ALANINE	52
ALARM	74
ALARM PROJECT	7
ALARMS	0

↓

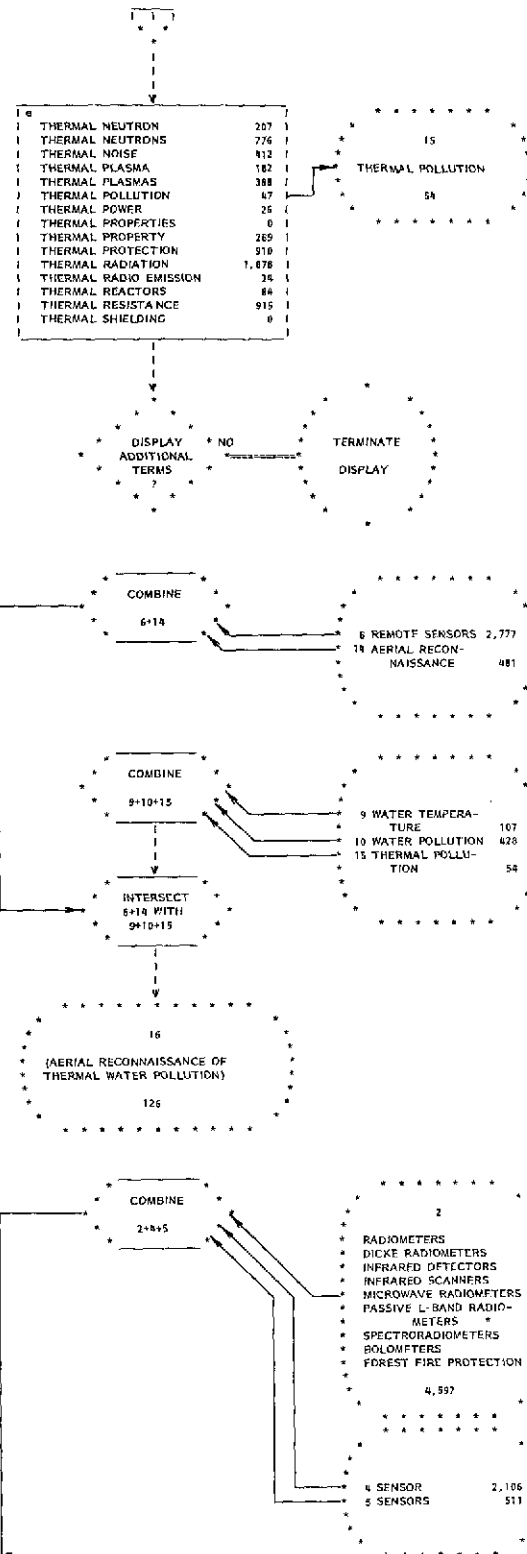
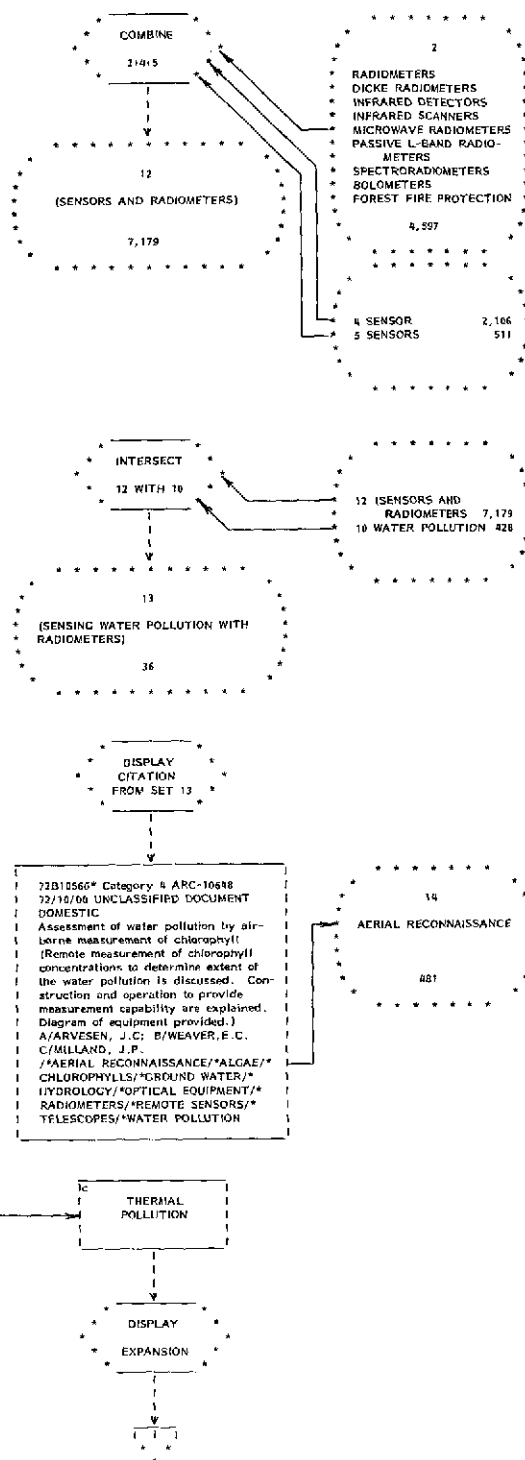
DISPLAY
LIMIT
REACHED



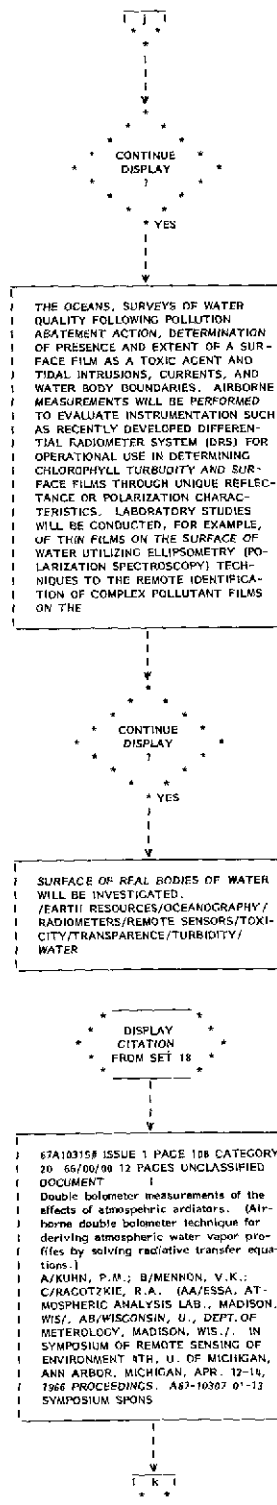
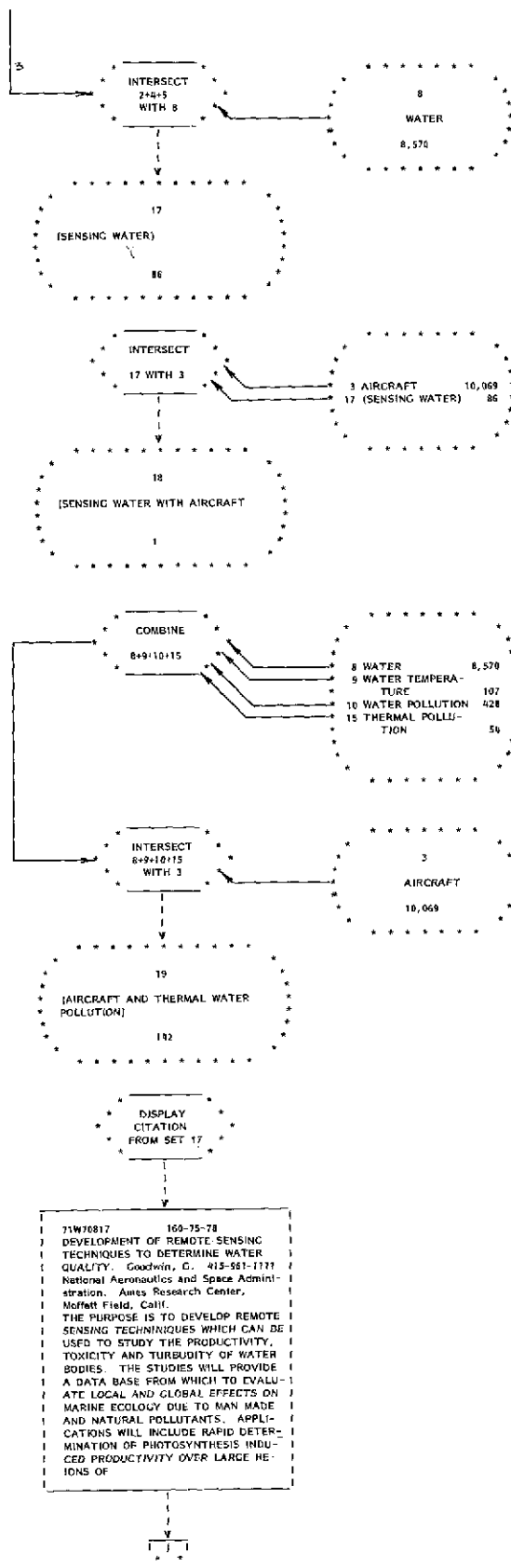


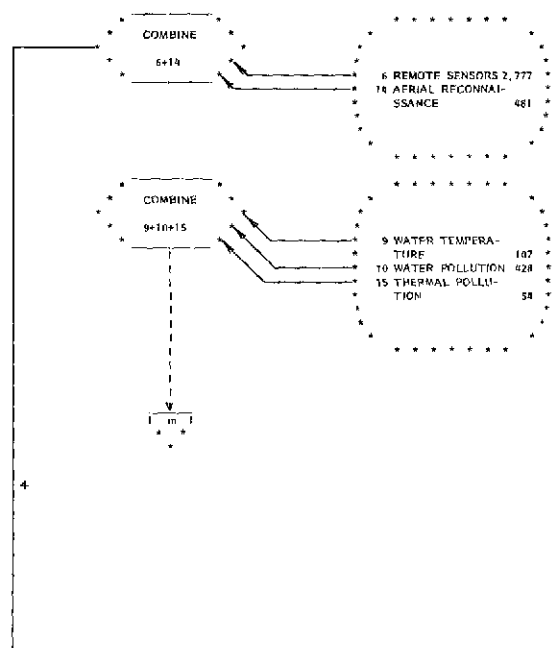
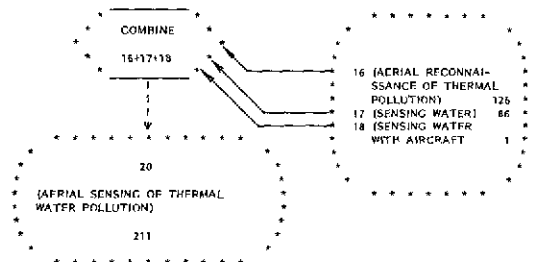
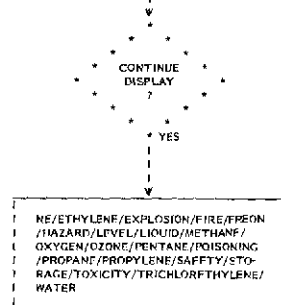
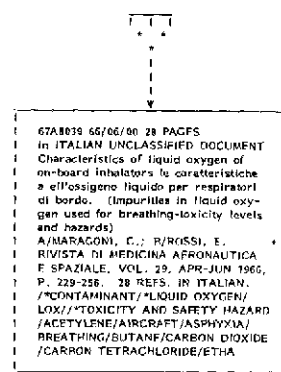
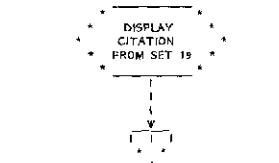
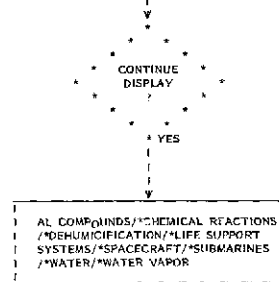
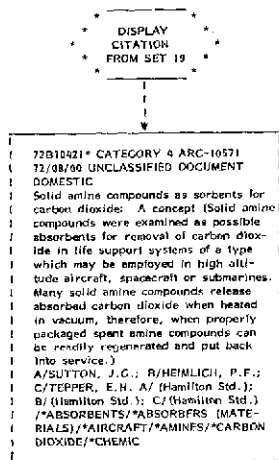
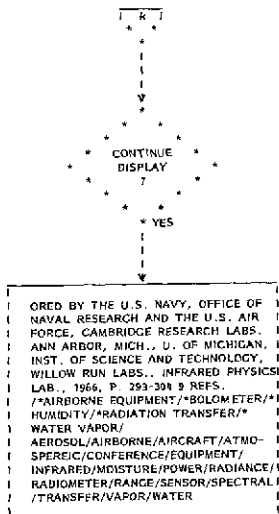


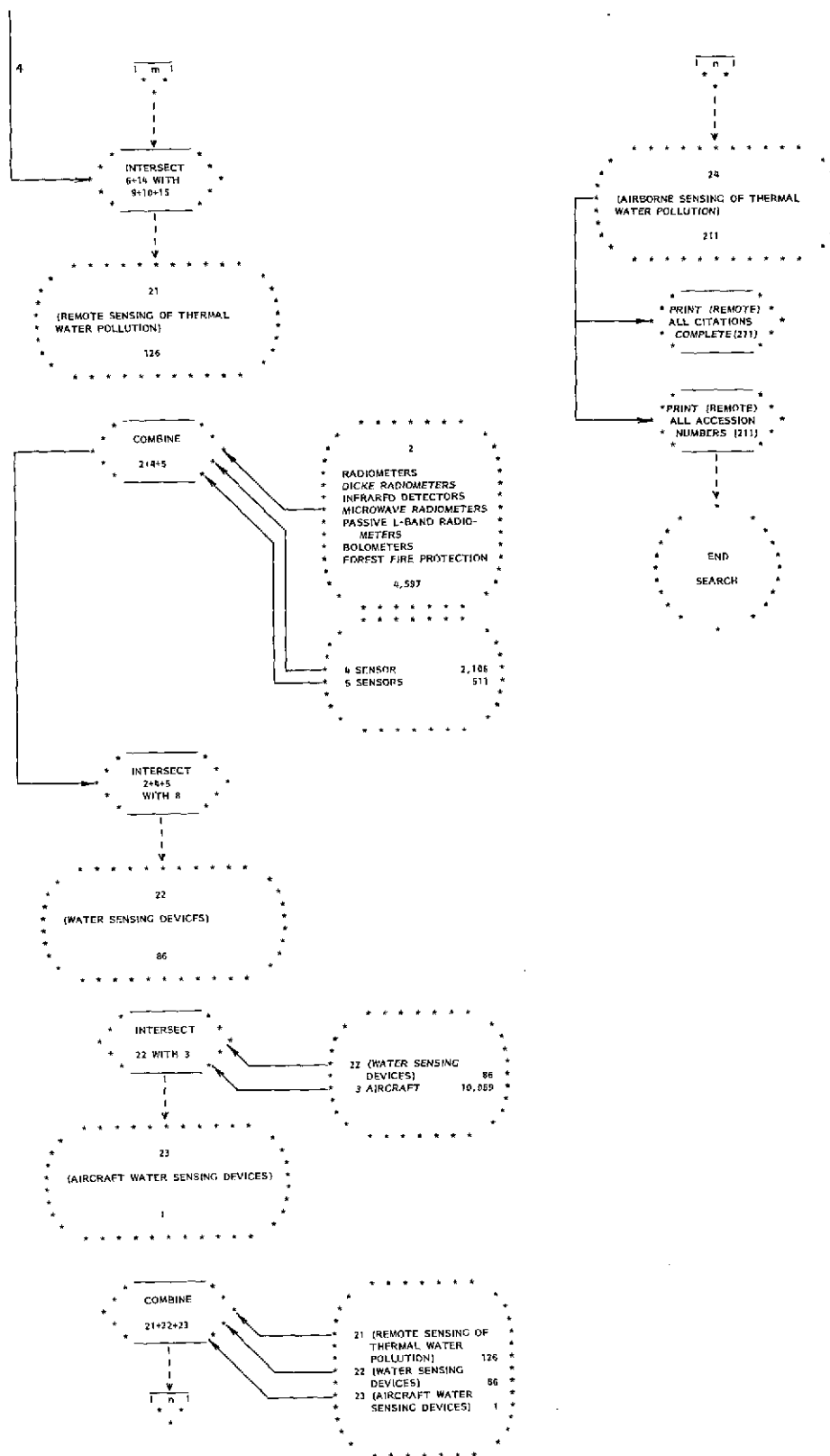
2



3







LOGICAL PROCESSES EMPLOYED BY SEARCHER TRP

Searcher TRP is a graduate electrical engineer with many years experience in component and device manufacturing industry. At NC/STRC his specialization has been in biomedical equipment. He is familiar with the NASA file and with RECON.

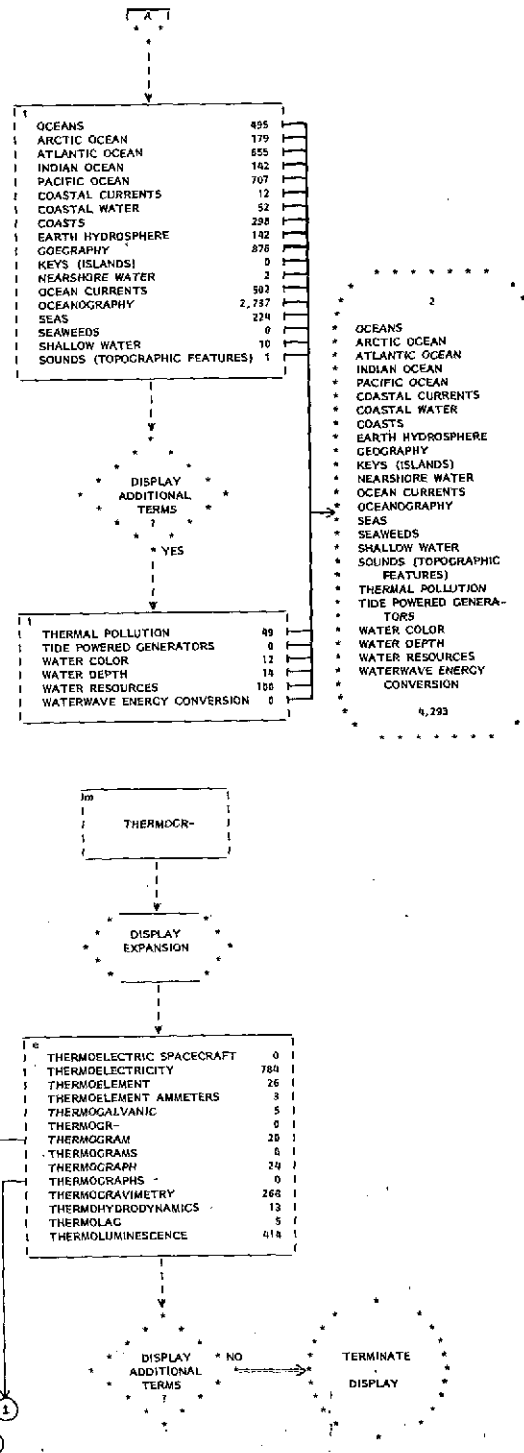
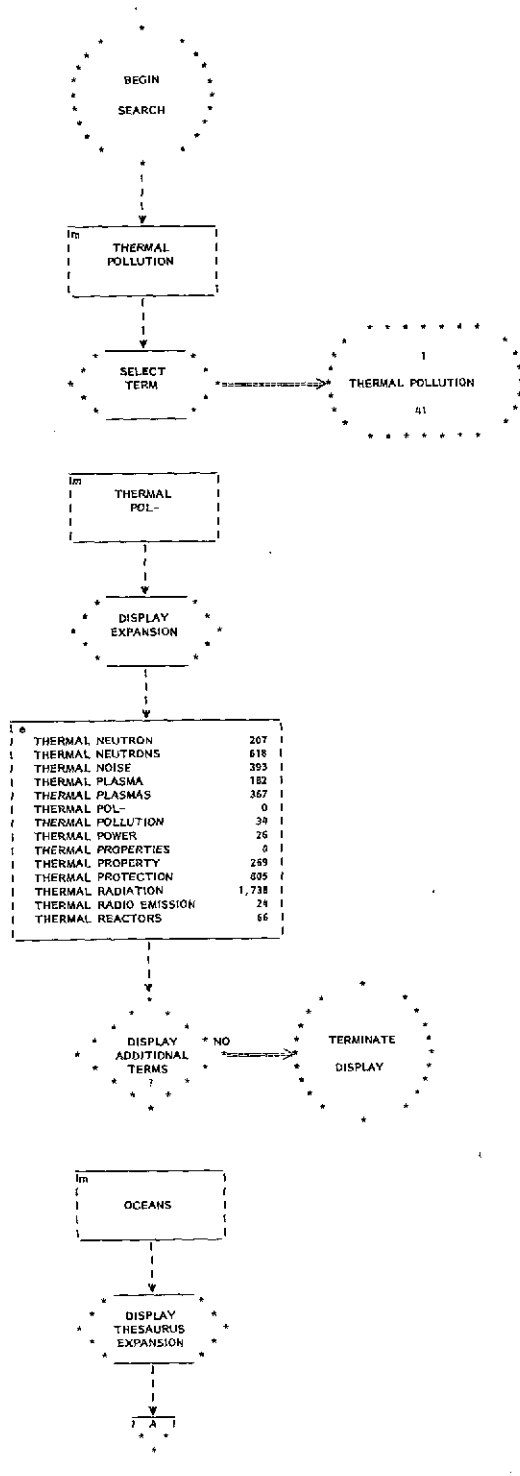
This searcher began his search, Figure 3, by selecting a term associated with the central thrust of the question, "thermal pollution." He then performed a dictionary expansion on the term to see if there was related material of interest, but found none. He then performed a thesaurus expansion on "oceans," and apparently with the idea that since oceans are larger than lakes more reports and papers deal with them. Much of the results, however, would be applicable in either case. From this expansion he selected 24 terms with a total of 4,293 postings to be set 2.

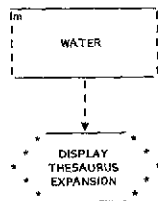
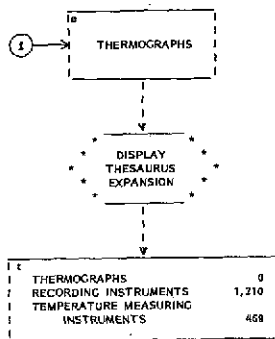
Following this, he asked for a dictionary expansion on the root "thermogr" which relates to the practice of recording temperature measurements. The expansion suggested to him that he make a thesaurus expansion on "thermographs." He decided against using any terms from this expansion, however. From a thesaurus expansion on water he selected 7 terms from which to create set 3. He then moved on to request a dictionary expansion on "remote s-" from which he decided to examine a thesaurus expansion on "remote sensors." This, however, did not seem to be particularly useful.

He next called for a thesaurus expansion on "radiometers" from which he selected 11 terms in order to create set 4 with 3,300 postings. At this point he also decided to select "thermograms" as set 5 and "remote sensors" as set 6, probably because he felt that the terms he had now seen under radiometers were no more pertinent than "thermographs" and "remote sensors." He then intersected the union of terms 4 and 6 with the union of terms 1, 2, 3, and 5. This produced set 7 with 626 hits, a respectable body of documents upon which one could probably perform further sorting advantageously. He attempted this by removing ocean radiometry (the intersection of set 4 with the union of sets 2 and 3) from set 7. This left him with 438 hits (set 9). Examination of the first two citations in set 9 indicated that this set was not highly responsive to the question.

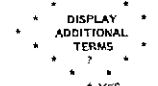
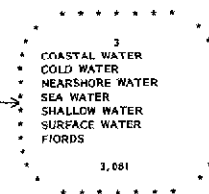
The searcher then felt that he should probably attack the problem along the lines of "temperature measurement." This expansion was not fruitful so he fell back to the base term "temperature." From this expansion he selected 14 terms with 43,078 postings as set 10 and another 14 terms with 6,332 postings as set 11. He had some second thoughts, however, about operating on such a large group of postings and recreated a display centered about "temperature field" from which he selected what he considered to be the most pertinent terms (8 with 13,146 postings) in order to form set 13. He then intersected this with set 6 (remote sensors) and added to this result the items deleted from set 9 dealing with radiometry of oceans (set 8). The resulting set of 319 documents he called his final hit list. Of the 319 hits, 97 or 30.4% were relevant. These 97 constituted 25% of the relevant documents found by all searches. Fourteen of the 97 relevant hits were unique to this search.

FIGURE 3

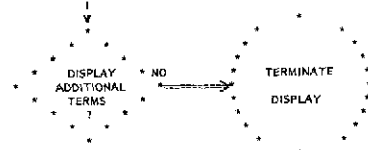
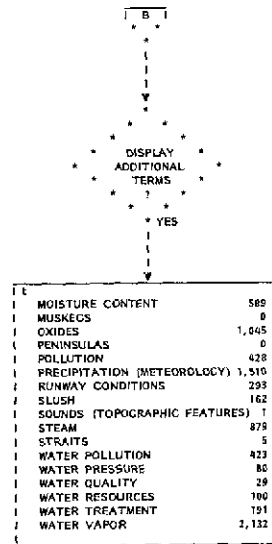




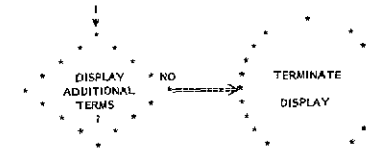
WATER	10,485
COASTAL WATER	32
COLD WATER	10
GROUND WATER	61
HEAVY WATER	195
NEARSHORE WATER	2
POLYMER WATER	11
POTABLE WATER	121
SEA WATER	1,321
SHALLOW WATER	10
SPRINGS (WATER)	0
SURFACE WATER	0
VADOSE WATER	11
AQUIFERS	0
ARROYOS	2
BAY ICE	1
BODY FLUIDS	181
CAVITATION FLOW	977
FJORDS	1

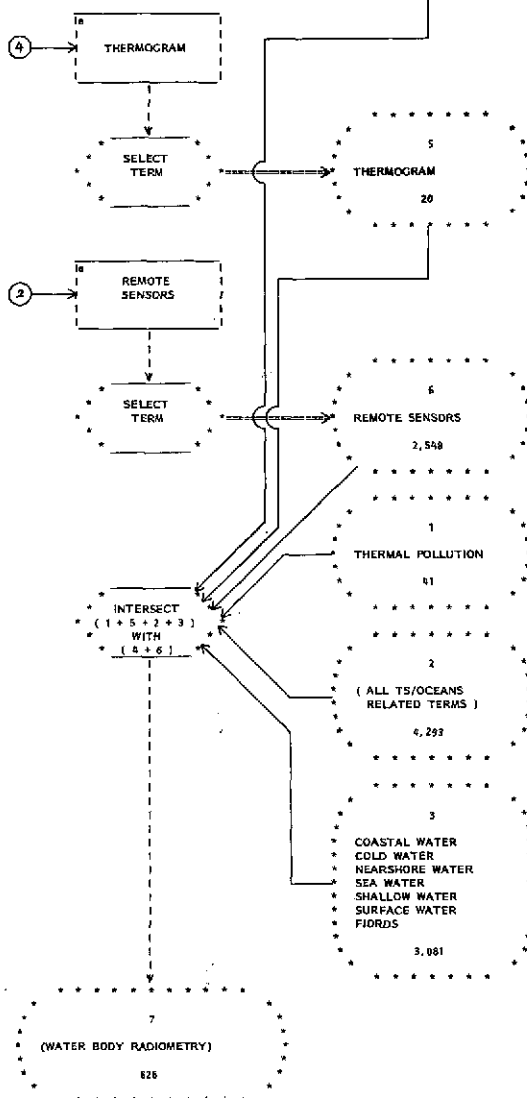
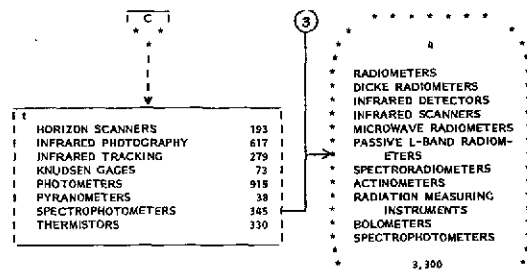
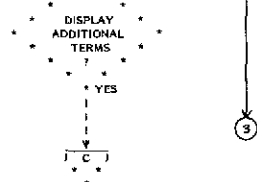
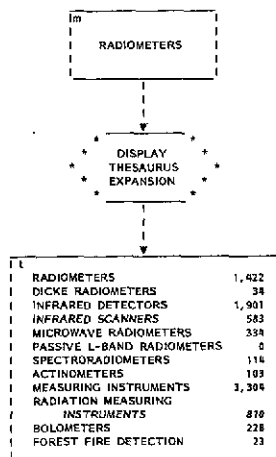
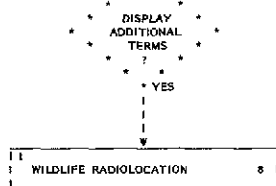
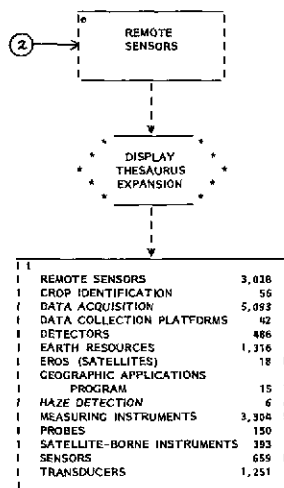


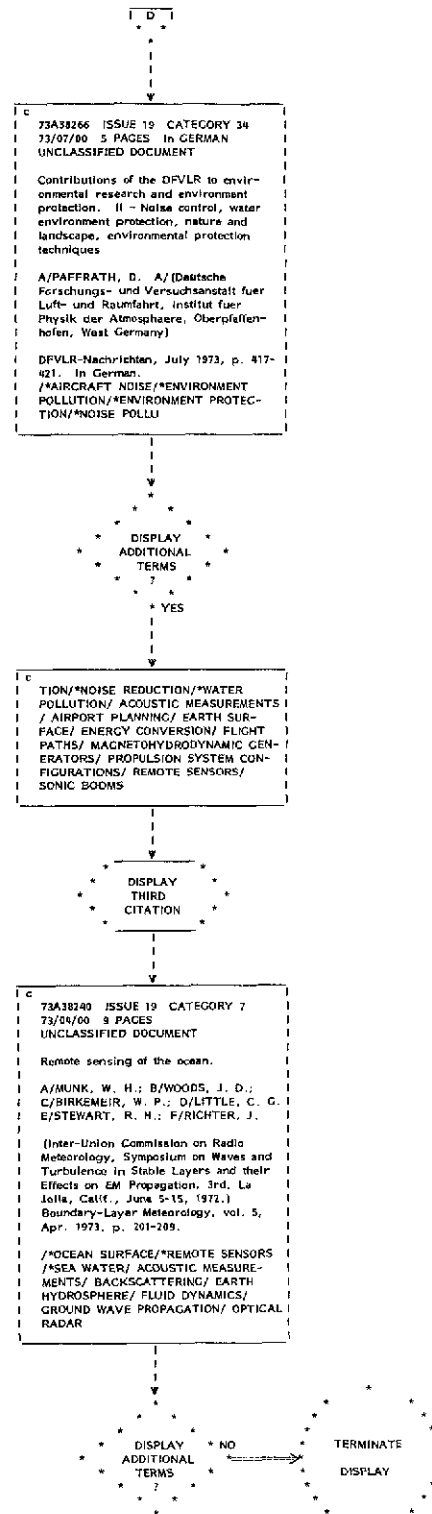
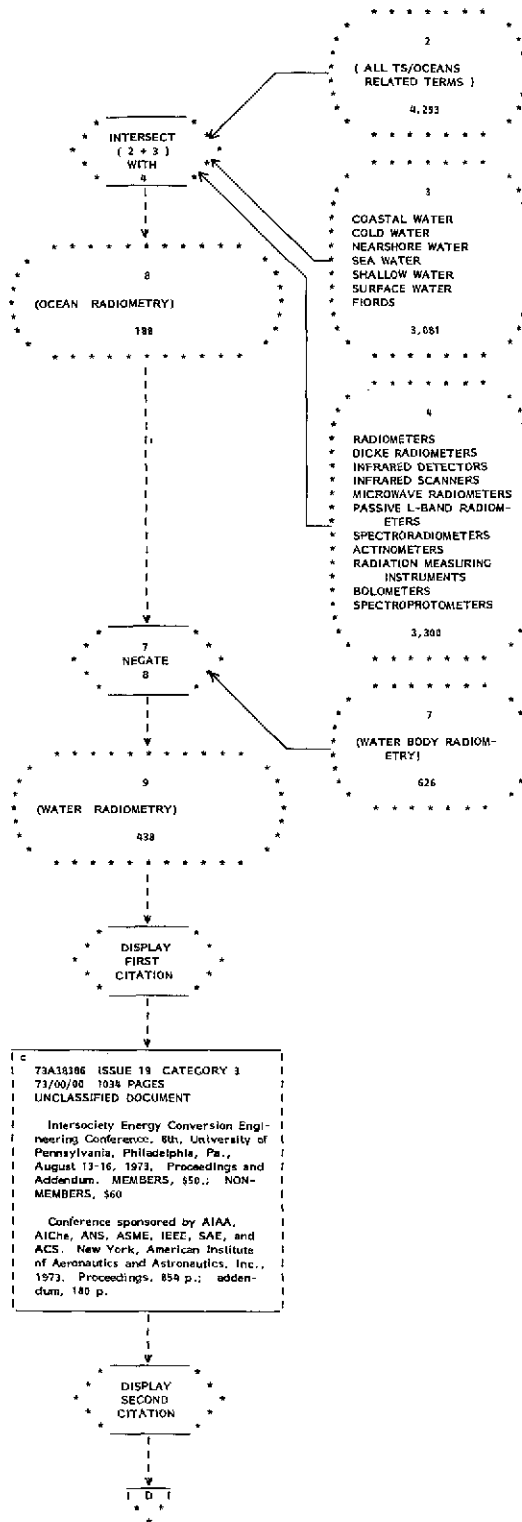
HUMIDITY	2,792
HYDRATES	127
HYDRAULICS	643
HYDRODYNAMICS	4,083
HYDROGEN COMPOUNDS	182
HYDROLOGY	746
HYDROMECHANICS	116
HYDROSTATICS	505
ICE	1,386
ISTHMOUSES	0
LAKE ICE	8
LATERITES	0
LIFE SUPPORT SYSTEMS	1,390
LIMNOLOGY	78
LIQUIDS	707
MODERATORS	167
MOISTURE	1,266



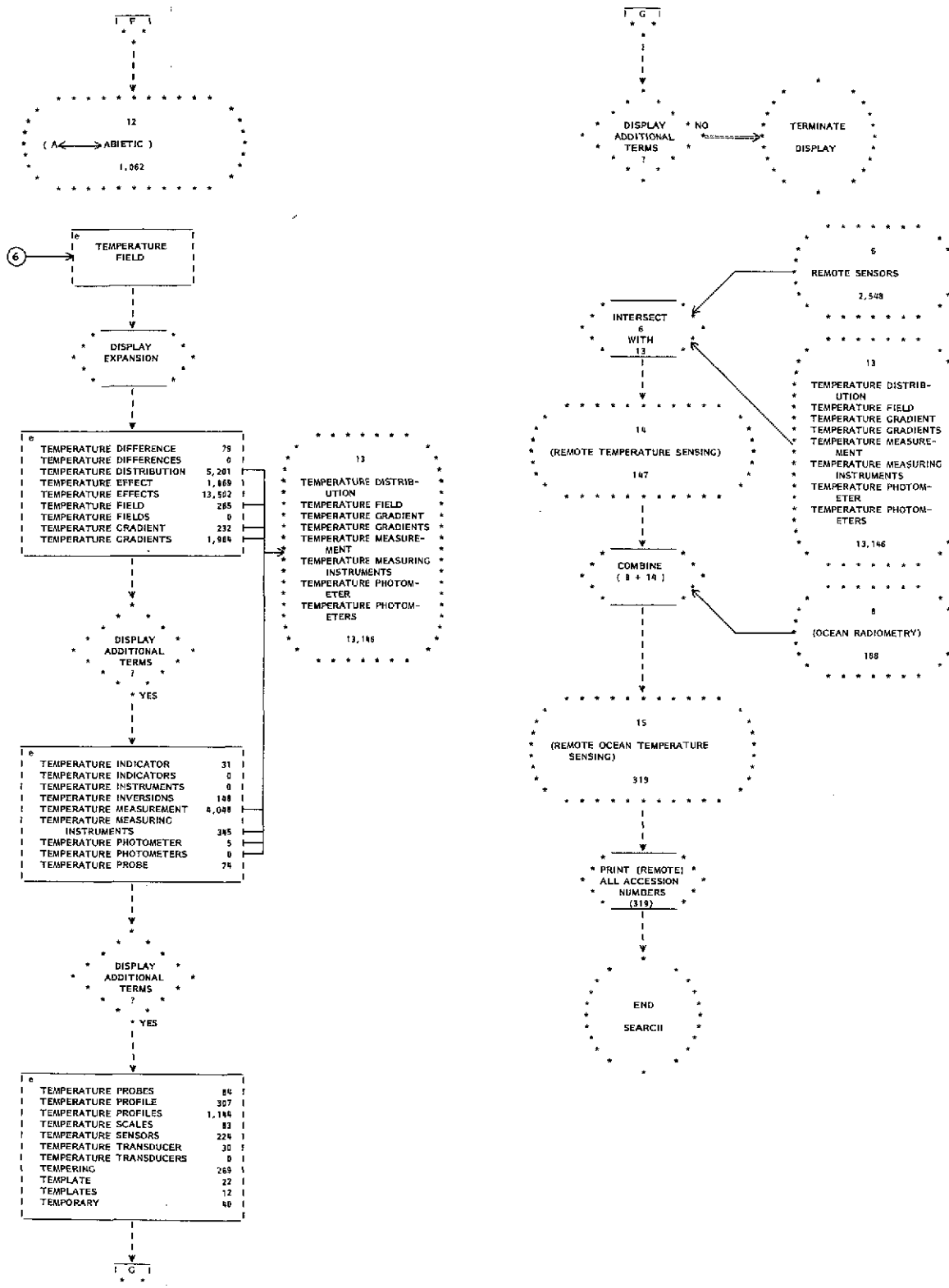
REMOTE	554
REMOTE CONSOLES	48
REMOTE CONTROL	684
REMOTE HANDLING	97
REMOTE REGIONS	4
REMOTE S-	0
REMOTE SENSORS	2,351
REMOTELY PILOTED VEHICLES	2
REMOVAL	502
REMS	0
RENAL	53
RENAL CALCULI	0
RENAL FUNCTION	146
RENDEZVOUS	637
RENDEZVOUS GUIDANCE	93
RENDEZVOUS GUIDANCE SYSTEM	80











LOGICAL PROCESSES EMPLOYED BY SEARCHER SRW

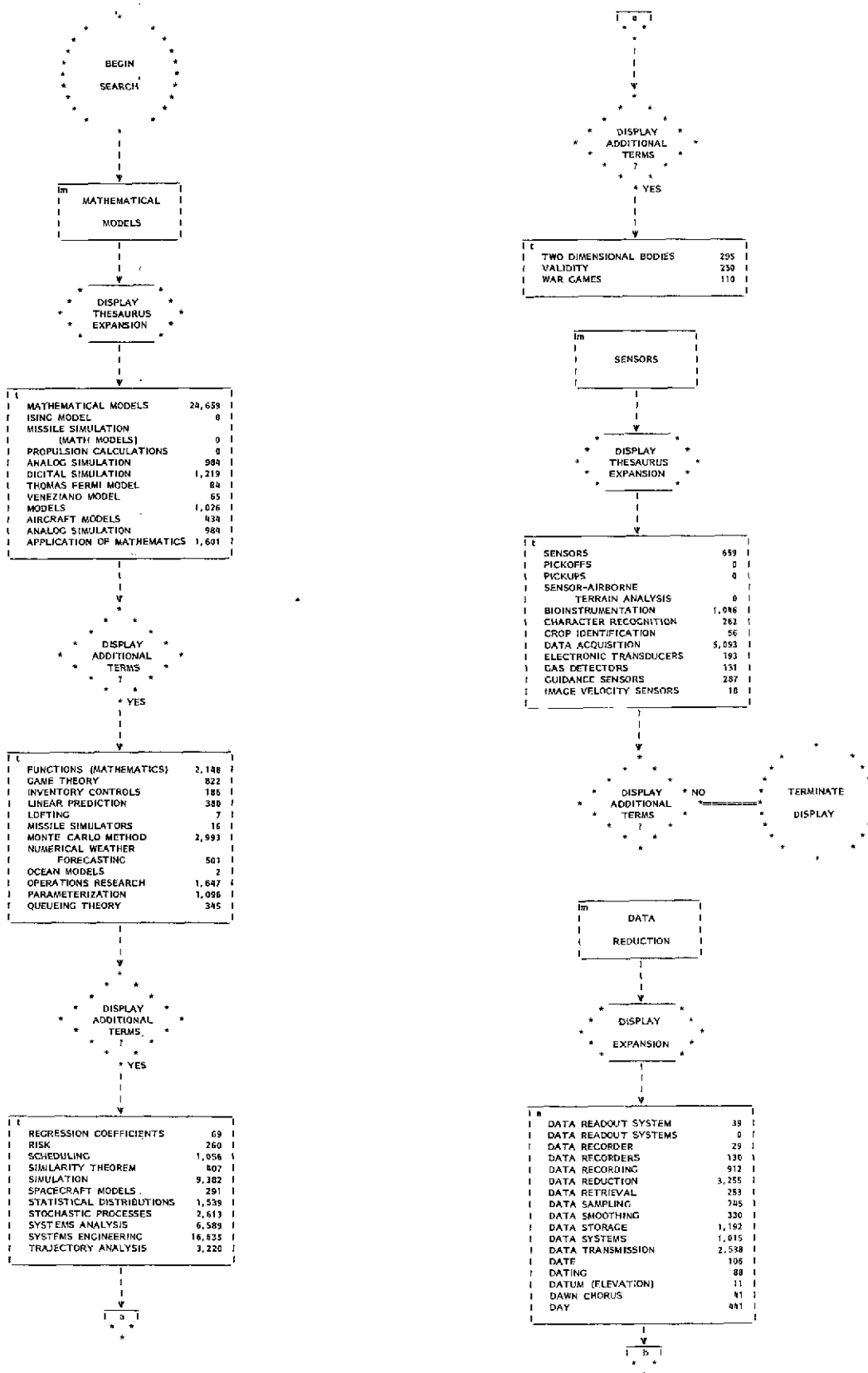
Searcher SRW is a graduate in English Education and is employed by NC/STRC in providing search service primarily to the education community and in general marketing. He is familiar with search techniques and philosophy and, to some extent, with the NASA file. He has no formal background, however.

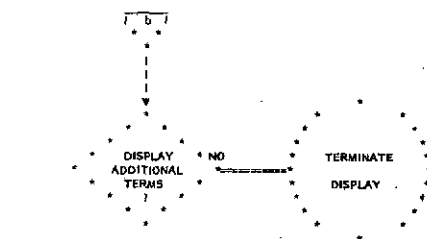
He began his search, Figure 4, with a thesaurus expansion on mathematical models. This concept, of course, is related to the interpretation of the sensor output. It is a bit strange, however, to begin at this point. He continued the expansion to its end without finding any terms he felt would be suitable. He then tried the idea of sensors. The thesaurus expansion did not seem to him to be particularly fruitful and was terminated after the initial display. He then moved to the idea of data reduction. The dictionary expansion here did not appear to be useful nor did a subsequent expansion on "data processing."

After drawing a blank from this line of questioning, the searcher moved to the area of mapping. From the thesaurus expansion, he picked the generating term along with "surveys" as set 1. A thesaurus expansion on "lakes" yielded three terms, "lakes," "surface water," and "thermal pollution" from which he formed set 2. Since thermal pollution was precisely the thing whose measurement was sought, he also tried a dictionary expansion on that term. From this expansion he selected "thermal radiation" as set 3. He then combined sets 2 and 3 and intersected this union with set 1 to obtain 26 items (set 4) dealing with mapping of lakes and thermal pollution.

Searcher SRW then examined the dictionary expansion on "remote sensors," selecting only the generating term as set 6. One other expansion occurred to him: "temperature measurement." This became set 7. Combining sets 1, 5, 6, and 7, he intersected these with the union of sets 2 and 3. Note that this is an expanded version of set 4. The 172 hits obtained from set 8 were selected as the final hit list. Forty-eight of these were regarded as pertinent by the experiment director. Thus while 27.9% of the hits were relevant, only 12.4% of those identified in the collection as being pertinent were recovered. Fourteen of the pertinent hits (or almost 30%) were not recovered by any other search.

FIGURE 4





DATA
PROCESSING

DISPLAY
EXPANSION

DATA HANDLING SYSTEM	250
DATA HANDLING SYSTEMS	0
DATA LINK	76
DATA LINKS	520
DATA MANAGEMENT	158
DATA PROCESSING	6,475
EQUIPMENT	730
DATA PROCESSING TERMINALS	115
DATA PROCESSOR	75
DATA PROCESSORS	0
DATA READOUT SYSTEM	39
DATA READOUT SYSTEMS	0

DISPLAY
ADDITIONAL
TERMS
?

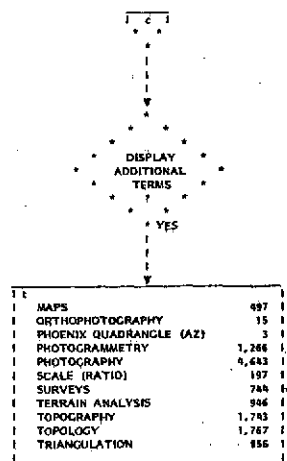
TERMINATE
DISPLAY

MAPPING

DISPLAY
THESAURUS
EXPANSION

MAPPING	3,858
CARTOGRAPHY	201
FLOX MAPPING	34
CADAstral MAPPING	17
ICE MAPPING	80
PHOTO MAPPING	48
SOIL MAPPING	154
THEMATIC MAPPING	16
ASTROGRAPHY	108
CONTOURS	473
DECLINATION	344
FIXED POINTS (MATHEMATICS)	247
FUNCTIONS (MATHEMATICS)	2,148
GEOGRAPHIC APPLICATIONS (PROGRAM)	13
GEOGRAPHY	876
HYPSOGRAPHY	0

1



MAPS	497
ORTHOPHOTOGRAPHY	15
PHOENIX QUADRANGLE (AZ)	3
PHOTOGRAMMETRY	1,266
PHOTOGRAPHY	4,643
SCALE (RATIO)	197
SURVEYS	744
TERRAIN ANALYSIS	966
TOPOGRAPHY	1,743
TOPOLOGY	1,787
TRIANGULATION	655

LAKES

DISPLAY
THESAURUS
EXPANSION

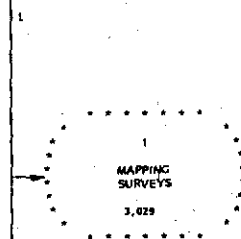
LAKES	195
BARRIER LAKES	0
CONSEQUENT LAKES	0
FINGER LAKES	0
FROZEN LAKES	0
MORAINAL LAKES	0
GREAT LAKES (NORTH AMERICA)	32
LAKE PONTCHARTRAIN (LA)	0
LAKE TEXOMA (OK-TX)	1
EARTH RESOURCES	1,316
AQUIFERS	0
BAYOUS	0
BEACHES	86
COASTS	298
EARTH HYDROSPHERE	142

DISPLAY
ADDITIONAL
TERMS
?

YES

EUTROPHICATION	4
KETTLES (GEOLOGY)	0
LAGOONS	1
LAKE CHAMPLAIN BASIN (NY-VT)	4
LAKE ICE	0
LIMNOLOGY	18
PLAYAS	16
PONDS	11
REGIONAL PLANNING	60
RESERVOIRS	98
RIVER BASINS	27
SPRINGS (WATER)	0
STRAITS	5
SURFACE WATER	0
THERMAL POLLUTION	49
WATER CIRCULATION	2
WATER COLOR	12
WATER DEPTH	12

2



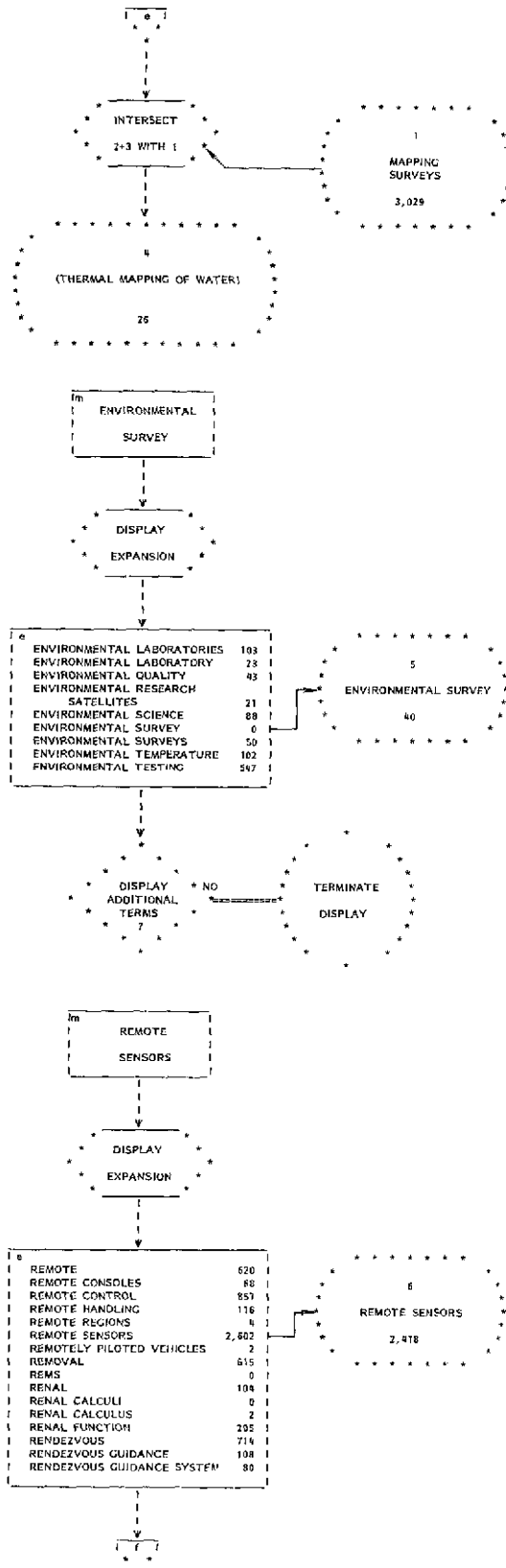
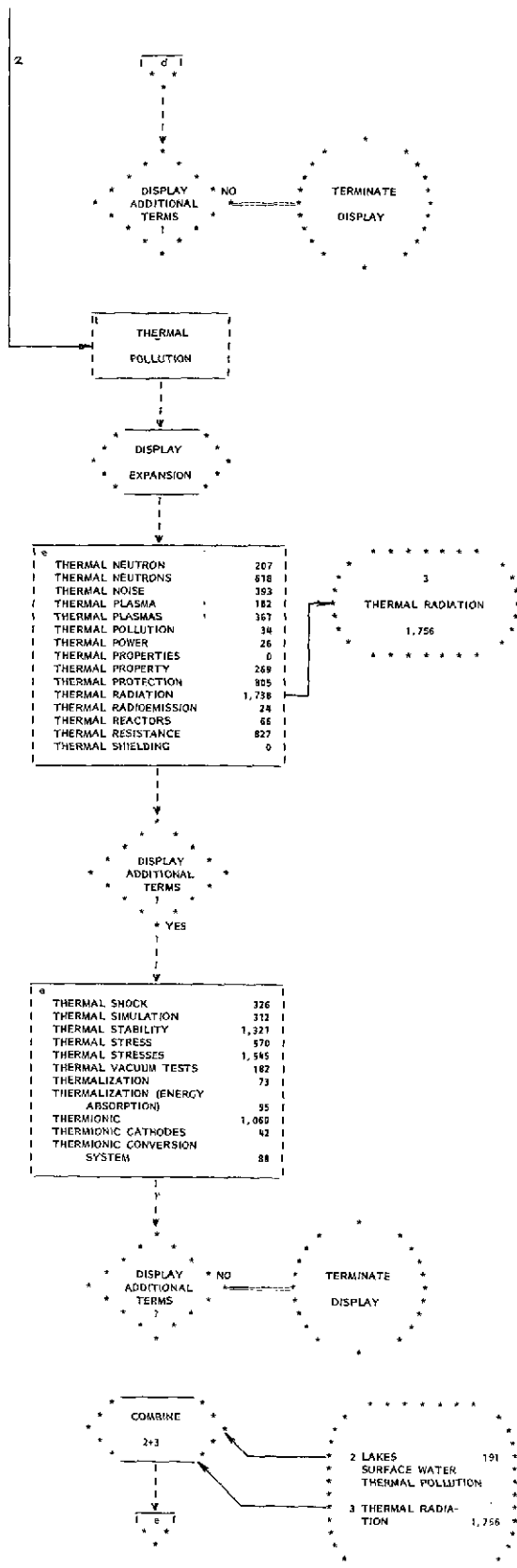
LAKES SURFACE WATER THERMAL POLLUTION	191
--	-----

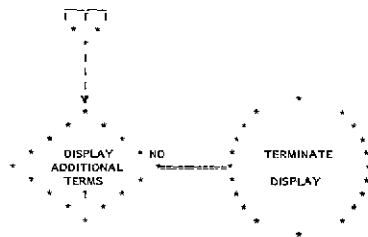
EUTROPHICATION	4
KETTLES (GEOLOGY)	0
LAGOONS	1
LAKE CHAMPLAIN BASIN (NY-VT)	4
LAKE ICE	0
LIMNOLOGY	18
PLAYAS	16
PONDS	11
REGIONAL PLANNING	60
RESERVOIRS	98
RIVER BASINS	27
SPRINGS (WATER)	0
STRAITS	5
SURFACE WATER	0
THERMAL POLLUTION	49
WATER CIRCULATION	2
WATER COLOR	12
WATER DEPTH	12

EUTROPHICATION	4
KETTLES (GEOLOGY)	0
LAGOONS	1
LAKE CHAMPLAIN BASIN (NY-VT)	4
LAKE ICE	0
LIMNOLOGY	18
PLAYAS	16
PONDS	11
REGIONAL PLANNING	60
RESERVOIRS	98
RIVER BASINS	27
SPRINGS (WATER)	0
STRAITS	5
SURFACE WATER	0
THERMAL POLLUTION	49
WATER CIRCULATION	2
WATER COLOR	12
WATER DEPTH	12

LAKES SURFACE WATER THERMAL POLLUTION	191
--	-----

EUTROPHICATION	4
KETTLES (GEOLOGY)	0
LAGOONS	1
LAKE CHAMPLAIN BASIN (NY-VT)	4
LAKE ICE	0
LIMNOLOGY	18
PLAYAS	16
PONDS	11
REGIONAL PLANNING	60
RESERVOIRS	98
RIVER BASINS	27
SPRINGS (WATER)	0
STRAITS	5
SURFACE WATER	0
THERMAL POLLUTION	49
WATER CIRCULATION	2
WATER COLOR	12
WATER DEPTH	12





TEMPERATURE
MEASURE

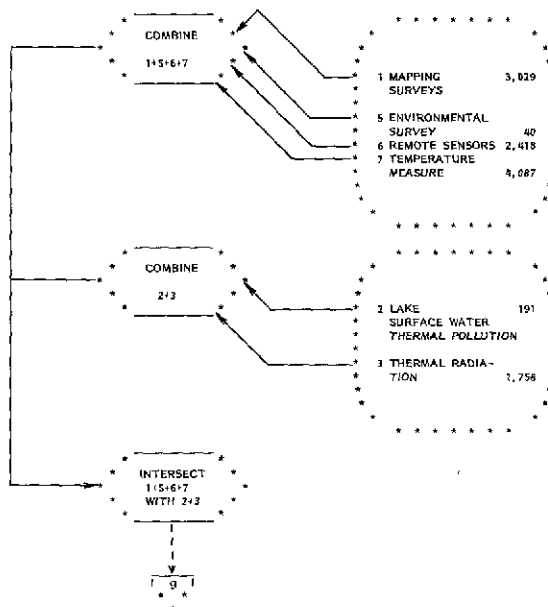
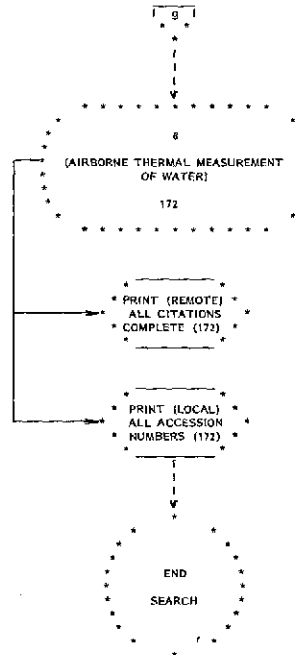
DISPLAY
EXPANSION

TEMPERATURE GRADIENTS	2,175
TEMPERATURE INDICATOR	32
TEMPERATURE INDICATORS	0
TEMPERATURE INSTRUMENTS	0
TEMPERATURE INVERSIONS	164
TEMPERATURE MEASURE	0
TEMPERATURE MEASUREMENT	4,380
TEMPERATURE MEASURING	
INSTRUMENTS	403
TEMPERATURE PHOTOMETER	5

TEMPERATURE MEASURE
4,087

DISPLAY
ADDITIONAL
TERMS
7

TERMINATE
DISPLAY



LOGICAL PROCESSES EMPLOYED BY SEARCHER AWL

This searcher is a graduate chemist with long research experience in the textile industry and recent professional experience in textile information retrieval. The NASA file and RECON are relatively unfamiliar to him.

He began his search (Figure 5) with a dictionary expansion on "water temperature." The disappointingly (to him) small number of postings and lack of related terms caused him to terminate the expansion after one display and ask for a thesaurus expansion. What he saw then also did not appeal to him. It did, however, suggest another avenue of investigation (an expansion on "surface temperature"). Before pursuing that, however, he decided to determine the possibilities of the term "lake." The dictionary expansion yielded one set (set 1). Then departing from the practice of the other searchers, he asked for and printed locally thesaurus expansions on lakes, surface temperature, radiometer, dicke radiometers, and infrared detectors as well as a dictionary expansion of radiometer without creating any sets at the time of the individual displays. This fact is not evident from an examination of Figure 5.

From this printed list of terms, he then selected as set 2 "surface temperature." He intended to perform an intersection of 1 and 2 but performed a union instead. This became set 3. The intersection became set 4 but had only 3 terms. He then consulted the printout again and selected "lakes" as set 5. This he intersected with 2 and got six hits, hardly better than "lake*" "surface temperature." Again he consulted the term printout and selected "water temperature" as set 7. An intersection of set 5 with set 7 gave only one hit, however. These three tries at identifying very specific documents he recognized to be failures so he returned to the practice of displaying and printing terms.

A dictionary and a thesaurus expansion on water were first. From this he selected "water" as set 9. From the previous printout he selected "radiometer" as set 10. An intersection of these two items produced set 11 with 49 hits. He then consulted his printout and selected "temperature" as set 12. An intersection of this with set 11 produced 20 hits. He then returned to the printout and selected nine additional sets (14-22) "coastal water," "sea water," "great lakes (North America)," "beaches," "coasts," "ponds," "reservoirs," "river basins," and "water resources." By this time he had realized that if he were going to retrieve a large percentage of the pertinent documents in the file he would have to search through a large number of documents.

He continued making thesaurus expansions on water-related terms: "rivers," "coastal water," "sea water," "great lakes (North America)," "ponds," and "reservoirs." In the process he also created sets 23, 24, 25, 26, 27, and 28, "rivers," "estuaries," "streams," "oceans," "ponds," and "reservoirs." He also selected "river" (set 29) and "stream" (set 30). He then began combining terms putting together 9, 14, 15, 16, 17, 18, 19, 20, and 21 to make set 31; and 22, 23, 24, 25, 26, 27, 28, 29, and 30 to make set 32. In set 33 he combined sets 31 and 32. The total postings in this set were 10,964.

Again he consulted his printout of terms. From this he selected 10 additional sets dealing with radiometers. He summed them in set 44 to obtain 3,377 postings. He then tried intersecting set 44 with set 33 to yield set 45 with 151 postings. This set, he expected, would contain documents dealing with radiometric measurements of water. To see what output would be generated by another water term "lake," he intersected set 44 with set 1 and got only 1 hit. Set 5, "lakes," was also intersected with set 44 to give set 47 with 7 hits. He then added sets 45, 46, and 47 to give a total of 157 hits (set 48). Departing slightly from this line of attack, he intersected set 7 (water temperature) with set 44 and got 12 hits. These he then added to set 48 to get 164 hits (set 50). He then summed set 50 with the results of three previous intersections: lake*surface temperature, set 4; lakes*surface temperature, set 8. The resulting set, set 51, had 172 hits.

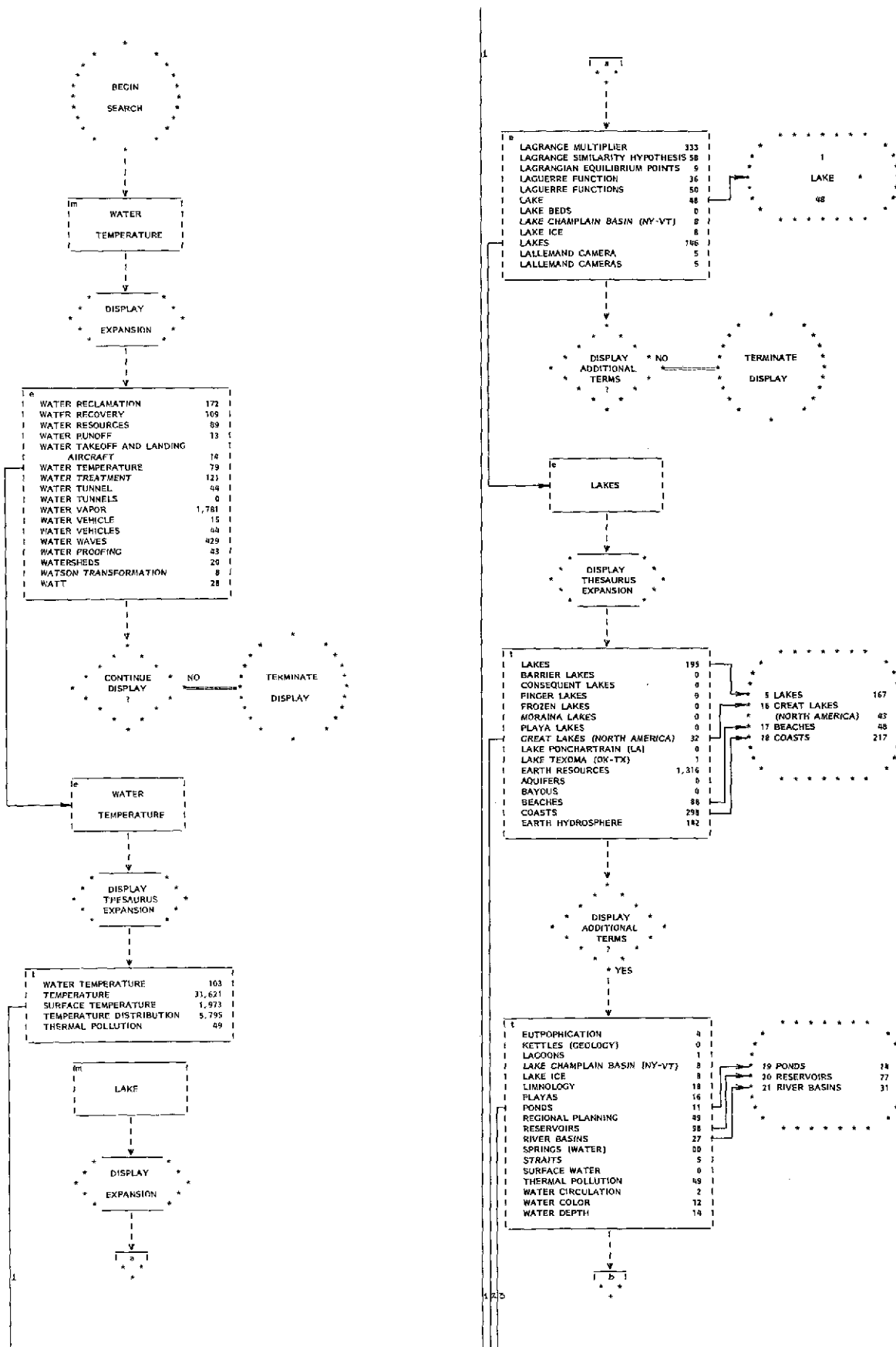
Again consulting his term printouts, he selected "measuring instruments" as set 52 and "radiation measuring instruments" as set 53. Departing from his previous practice, he selected as sets directly without first viewing as expansions several synonyms for that portion of the measuring instrument in contact with the phenomenon being observed: sensors (set 54), sensor (set 55), detectors (set 56), detector (set 57), scanners (set 58), scanner (set 59), probes (set 60), and probe (set 61). These 10 he summed to obtain set 62 with 12,471 postings.

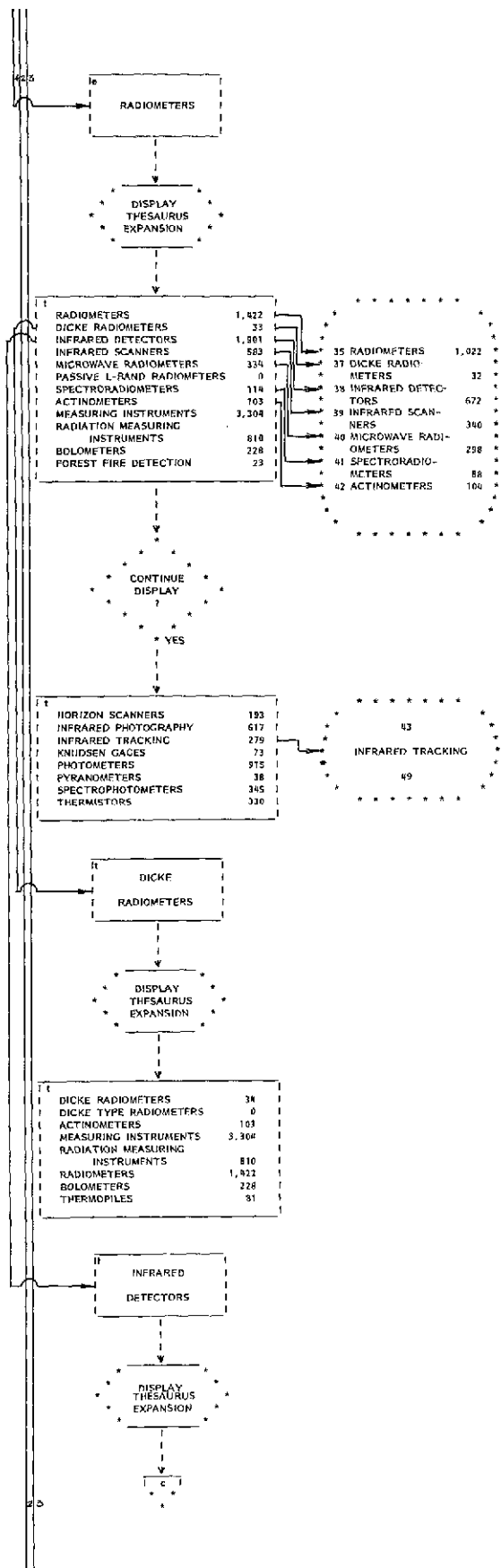
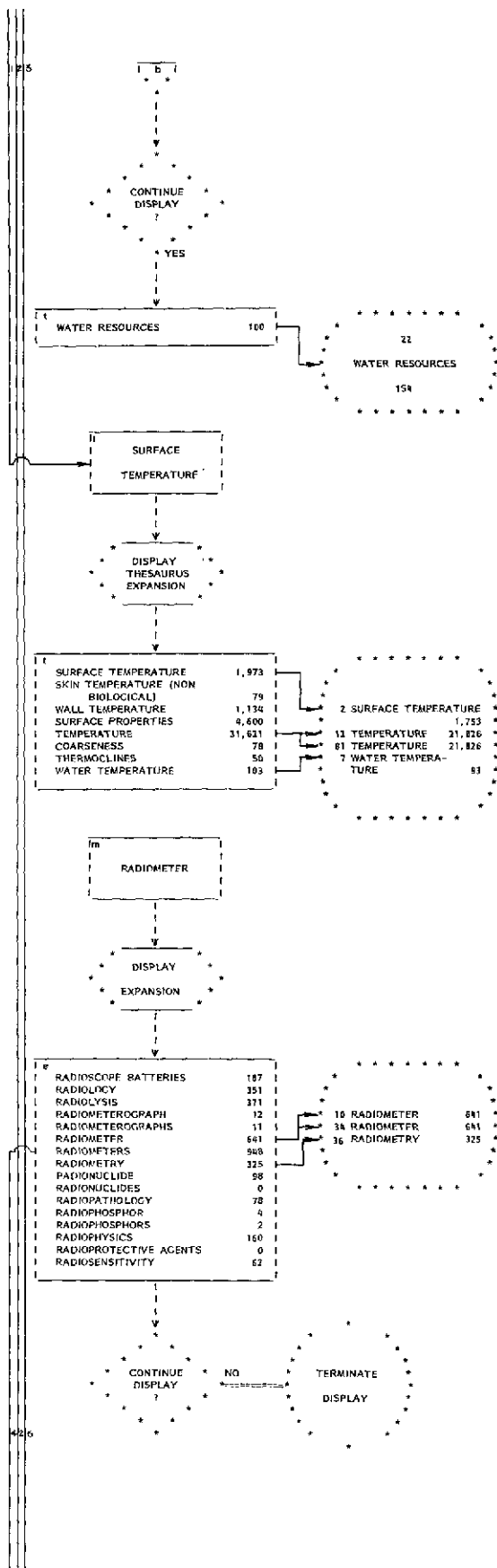
Having decided that to insure high recall one must examine a large number of documents, he added to set 33 sets 1 (lake), 5 (lakes), and 7 (water temperature) to create set 63. Although he had already used these sets previously, he apparently decided to broaden the search somewhat by including in the investigation, at least for the time being, possible relevant documents which may have been eliminated during previous intersections. Set 63 had 11,155 postings compared with 10,964 in set 33 alone. An intersection of set 62 with set 63 yielded 292 hits. These hits should be related to measuring things about water, though not specifically the radiation emitted. He added these 292 hits to the 172 he had obtained previously (set 51) which he felt were reasonably specific with regard to radiation measurement of bodies of water, after first ascertaining that intersecting either set 51 or set 64 with set 12 (temperature) would give a very small output. He then elected to print the hit list for the 447 items in set 67. A disc overflow temporarily interrupted his search for the final aspect of the search: mathematical models for interpreting radiometer indications.

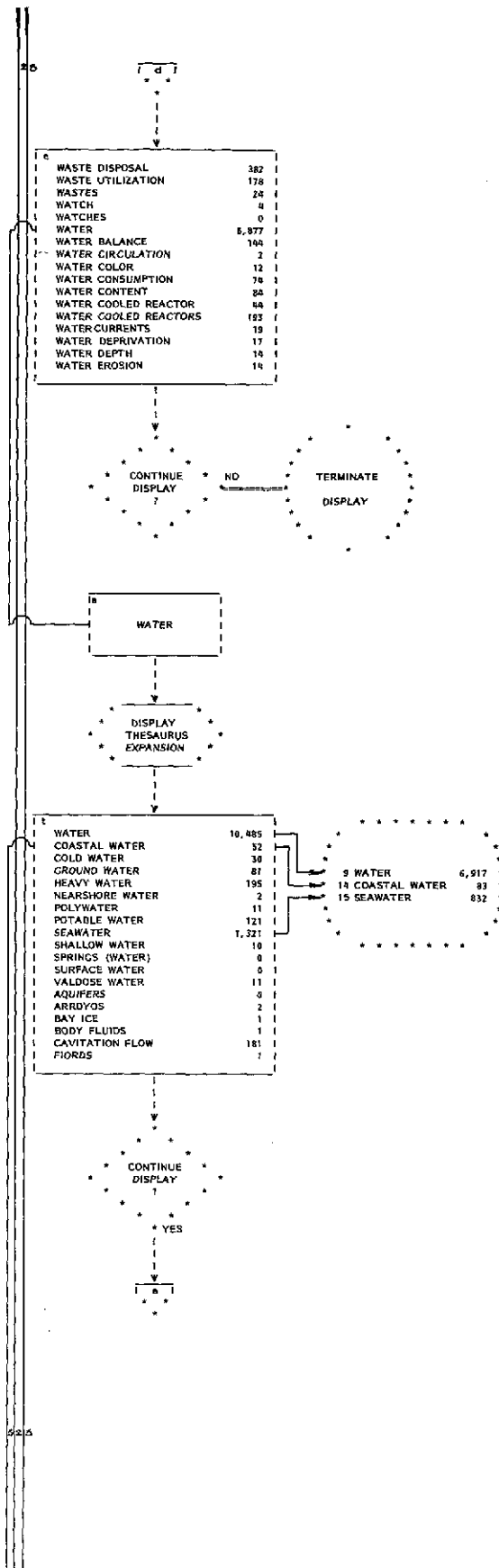
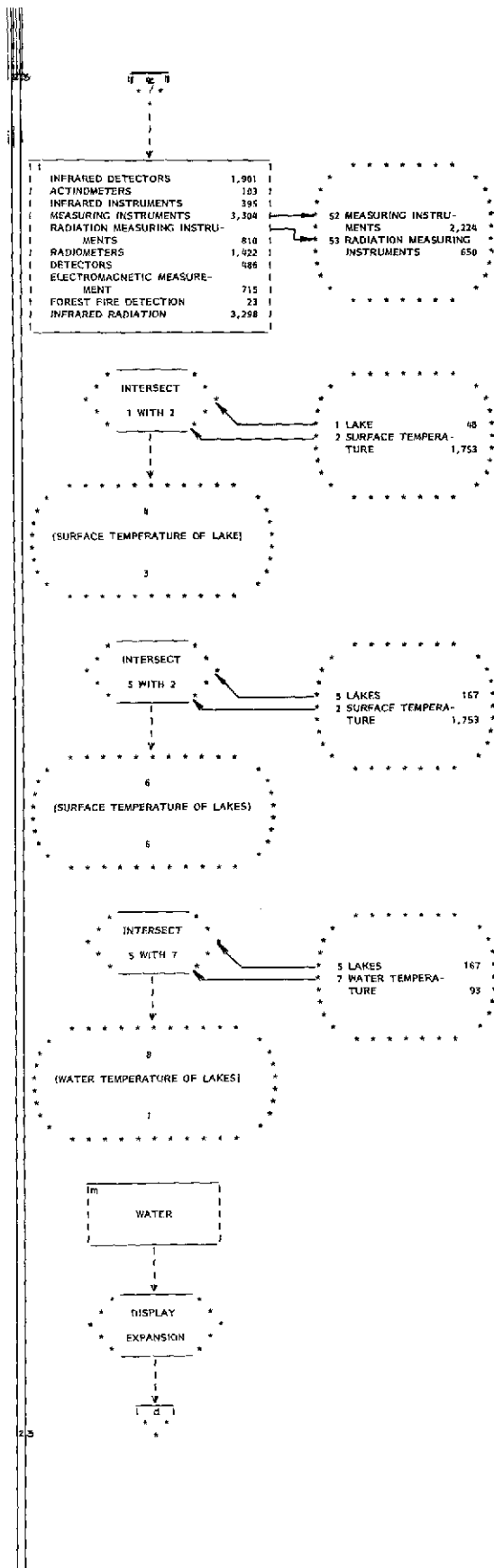
When he was able to resume his search, he performed expansions on "mathematical models," "data reduction," and "computer program." From these dictionary expansions he selected "mathematical logic" (set 68), "mathematical model" (set 69), "mathematical models" (set 72), "data reduction" (set 73), "computer program" (set 74), and "computer programs" (set 75). The numbers for sets 70 and 71 were unused. The six foregoing items were summed to create set 76 with 34,935 items. This was intersected with the union of "thermal radiation" and "infrared radiation" to yield set 80 with 223 postings. After seeing that an intersection of set 80 with the term "temperature" only produced 3 hits, searcher AWL asked that the items in set 80 be printed and added to those previously printed (set 67). These two groups he called his final hit list.

Of this total of 670 documents, 87 or 13% were judged to be relevant by the experiment director. This search found 22.4% of the relevant documents identified as being in the NASA collection. Ten of the 87 relevant documents were not found by any other search. It may be pointed out that although this search utilized a large number of terms and a large number of postings, comparatively few terms were in the area of temperature measurement, radiant heat flux measurement, thermal emission, heat transfer, etc. This apparently is the reason that this broad search failed to retrieve more relevant documents.

FIGURE 5







HUMIDITY	2,792
HYDRATES	127
HYDRAULICS	682
HYDRODYNAMICS	4,083
HYDROGEN COMPOUNDS	182
HYDROLOGY	746
HYDROMECHANICS	166
HYDROSTATICS	509
ICE	1,386
ISTHMOUSES	9
LAKE ICE	8
LATERITES	0
LIFE SUPPORT SYSTEMS	1,340
LIMNOLOGY	18
LIQUIDS	707
MODERATORS	167
MOISTURE	1,266

CONTINUE
DISPLAY
?

YES

MOISTURE CONTENT	588
MUSKEGS	0
OXIDES	1,085
PENINSULAS	0
POLLUTION	428
PRECIPITATION (METEOROLOGY)	1,510
RUNAWAY CONDITIONS	293
SLUSH	162
SOUNDS (TOPOGRAPHIC FEATURES)	1
STEAM	879
STRAITS	5
WATER POLLUTION	423
WATER PRESSURE	80
WATER QUALITY	29
WATER RESOURCES	100
WATER TREATMENT	191
WATER VAPOR	2,132

CONTINUE
DISPLAY
?

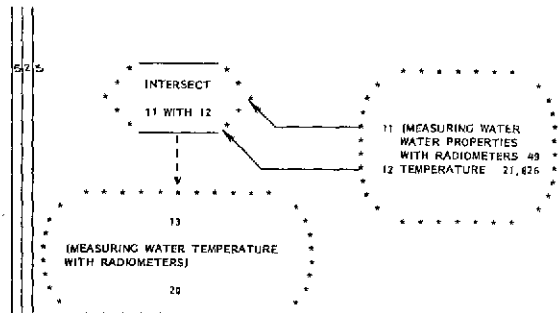
YES

WATERSHEDS	23
WHARVES	0
WINDPOWERED PUMPS	0

INTERSECT
9 WITH 10

9 WATER
10 RADIOMETER 6,917
641

11
(MEASURING WATER PROPERTIES
WITH RADIOMETERS)
49



RIVERS

DISPLAY
THESAURUS
EXPANSION

RIVERS	245
AMAZON REGION (SOUTH AMERICA)	0
EARTH RESOURCES	1,316
ALLUVIUM	2
BARITO RIVER BASIN (INDO- NESIA)	0
BAYDUS	0
CANYONS	0
DELAWARE RIVER BASIN (US)	0
DELTA	3
EROSION	1,907
ESTUARIES	64
MEANDERS	2
MISSISSIPPI DELTA (LA)	11
RAPIDS	0
RESERVOIRS	98

23 RIVERS 212
24 ESTUARIES 60

CONTINUE
DISPLAY
?

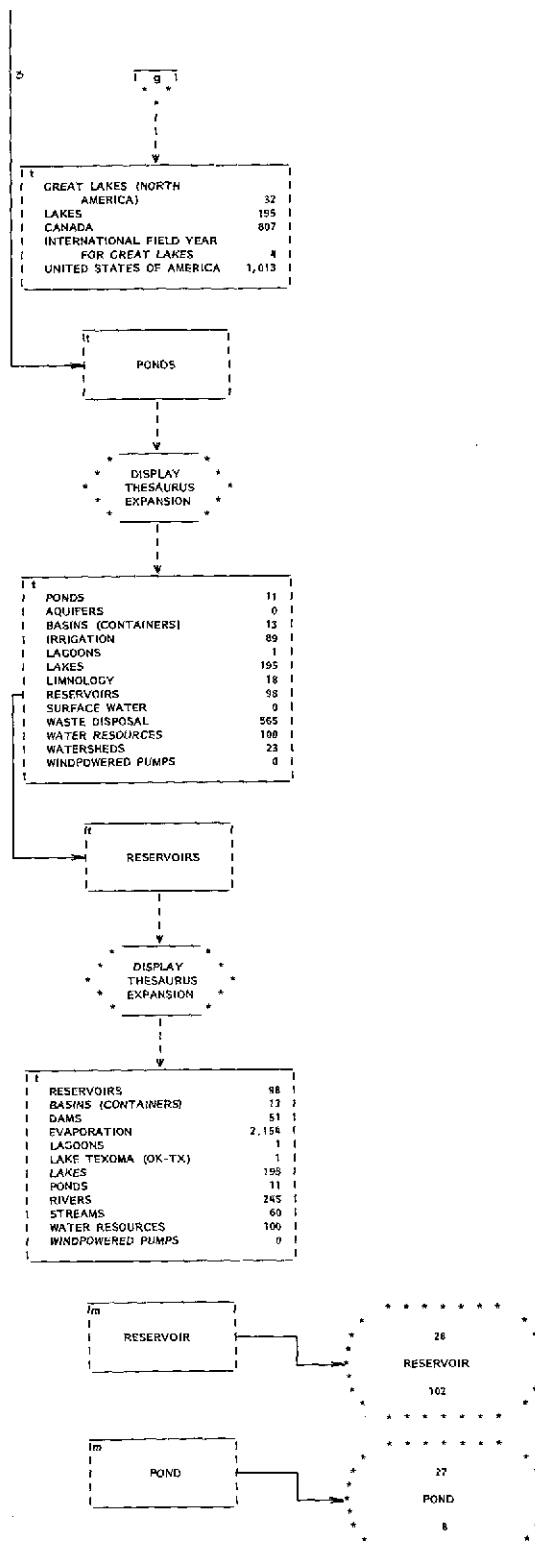
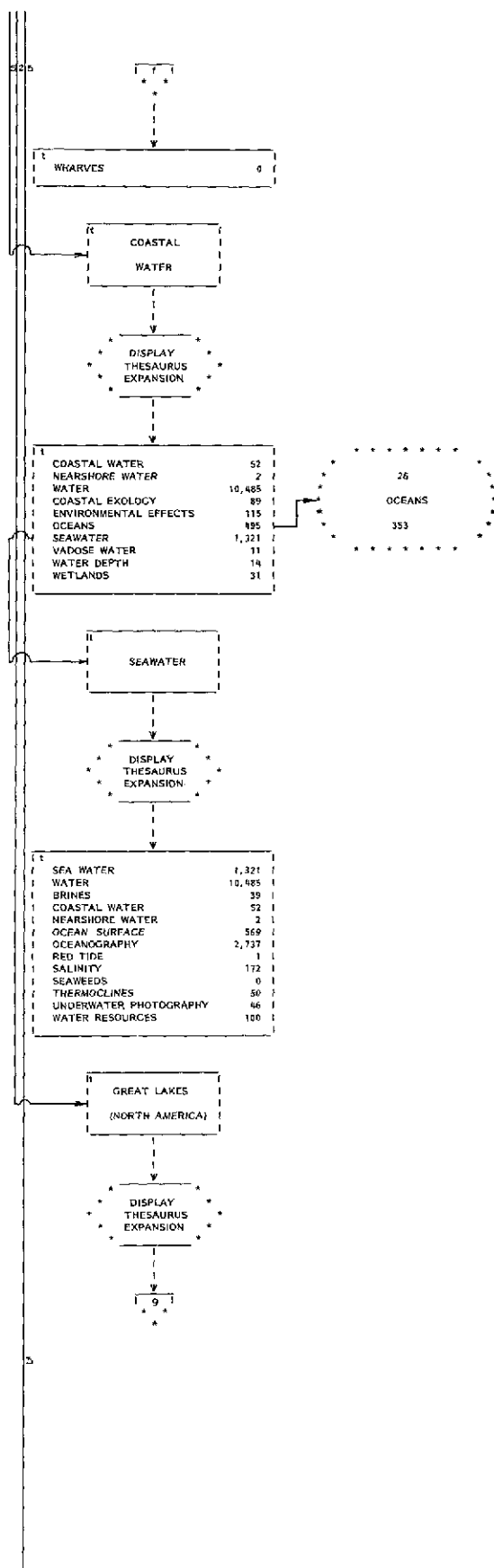
YES

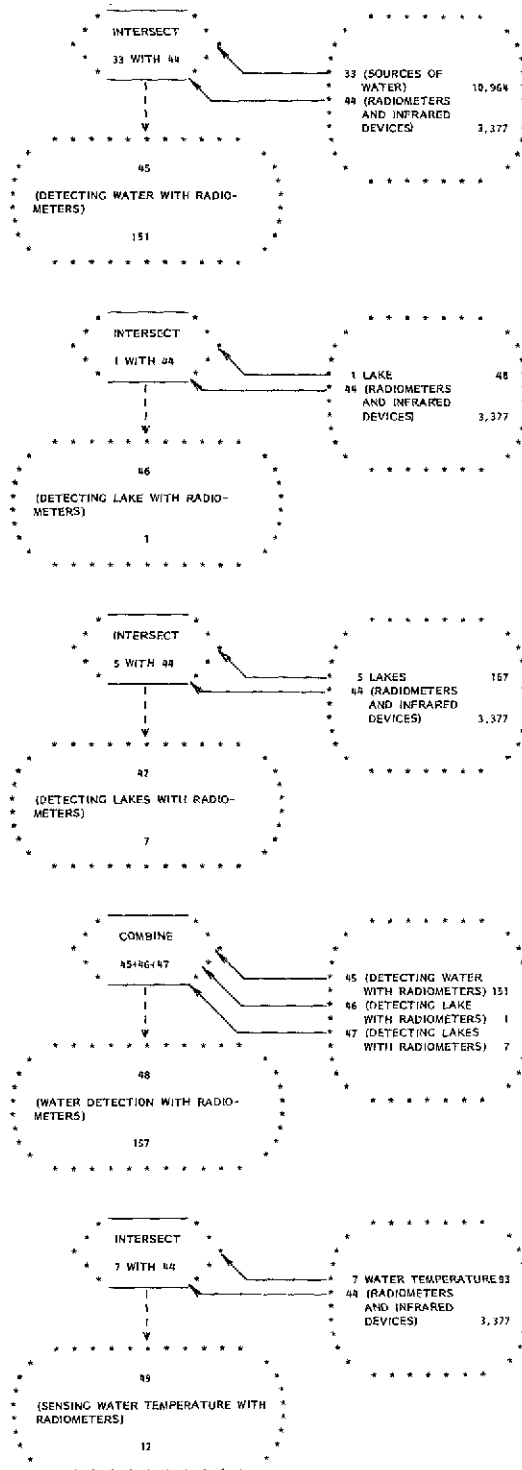
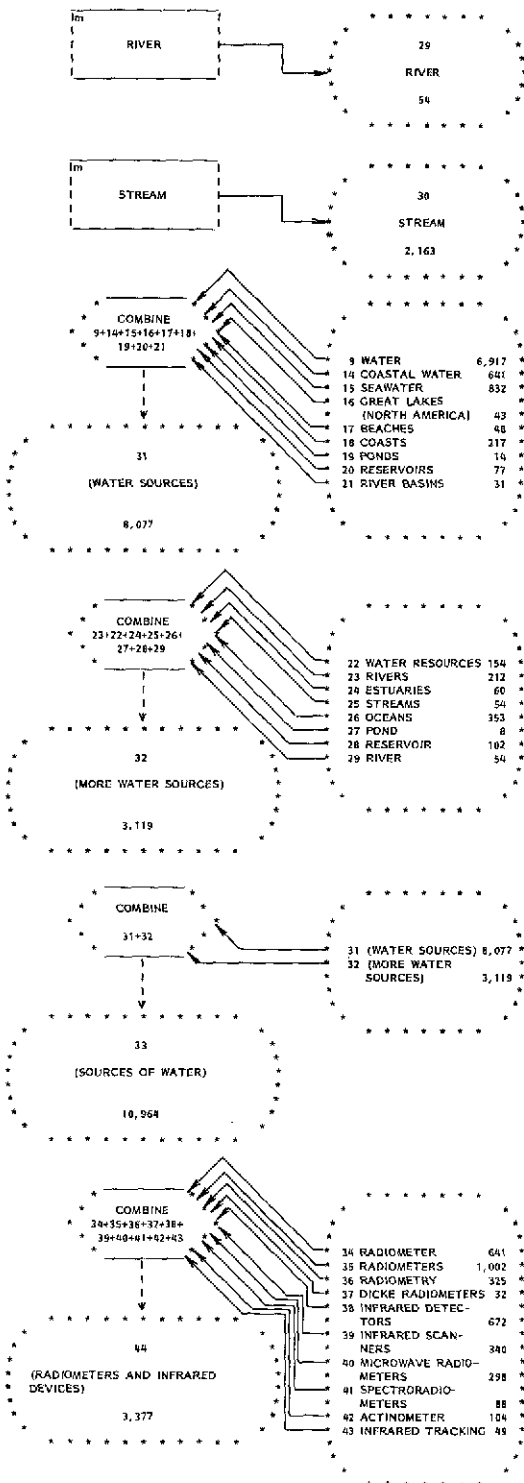
RHONE DELTA (FRANCE)	0
RIVER BASINS	27
SANTA RIVER BASIN (PERU)	0
SOUNDS (TOPOGRAPHIC FEATURES)	1
STREAMS	60
SURFACE WATER	0
SUSQUEHANNA RIVER BASIN (MD-NY-PA)	1
TRIBUTARIES	0
VALLEYS	66
WABASH RIVER BASIN (IL-IN OH)	4
WADIS	0
WATER COLOR	12
WATER RUNOFF	18
WATERSHEDS	23

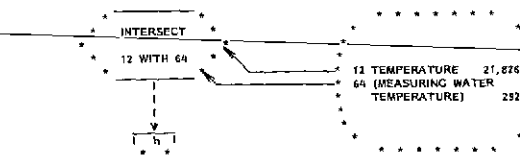
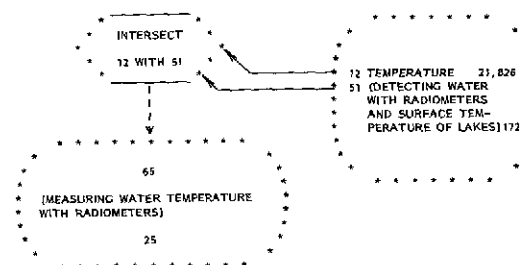
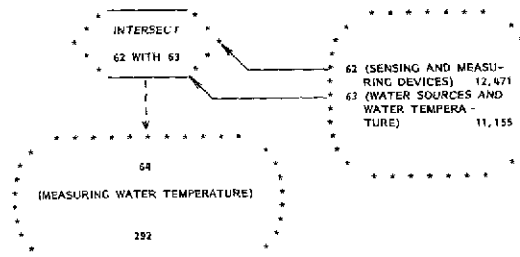
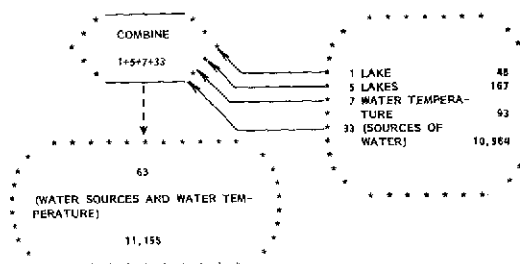
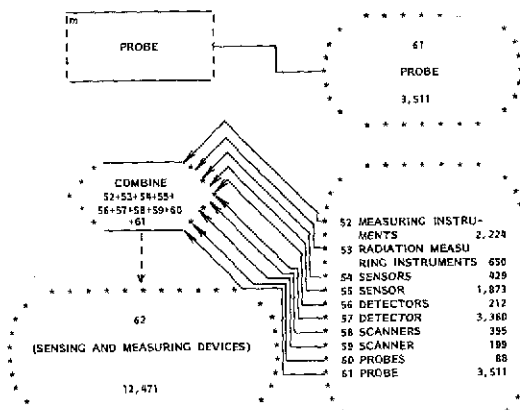
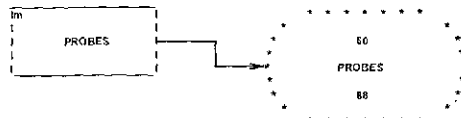
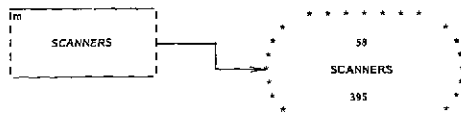
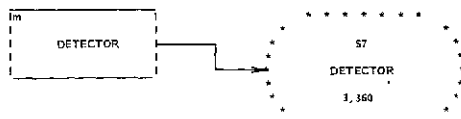
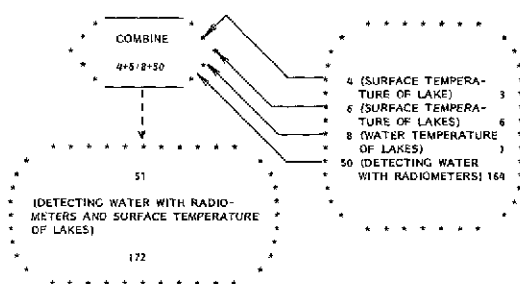
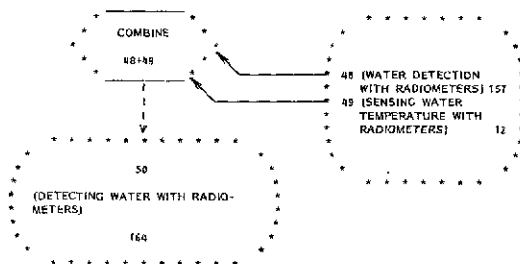
25
STREAMS 54

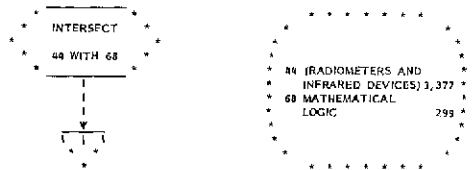
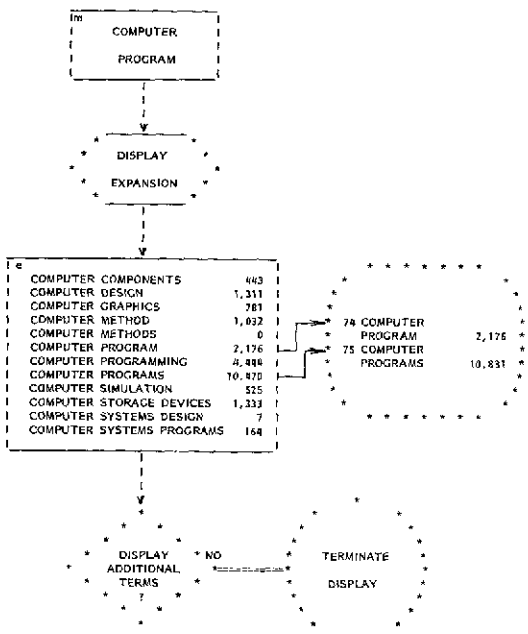
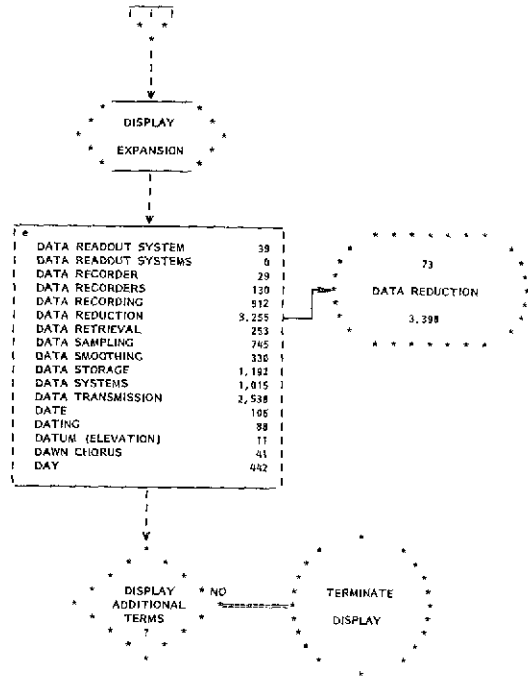
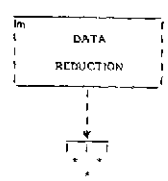
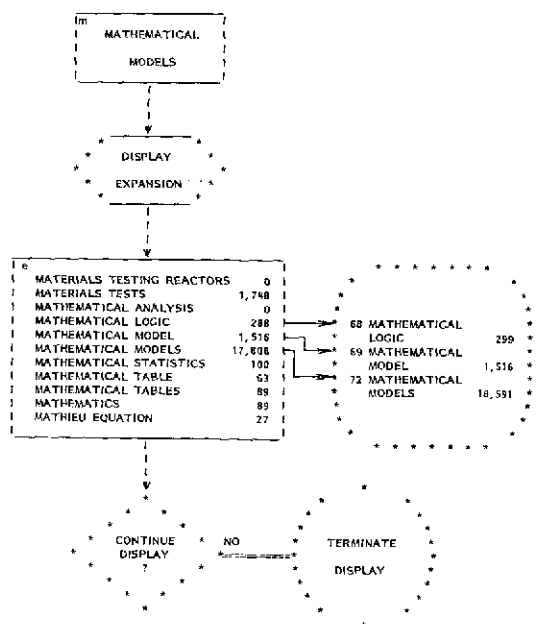
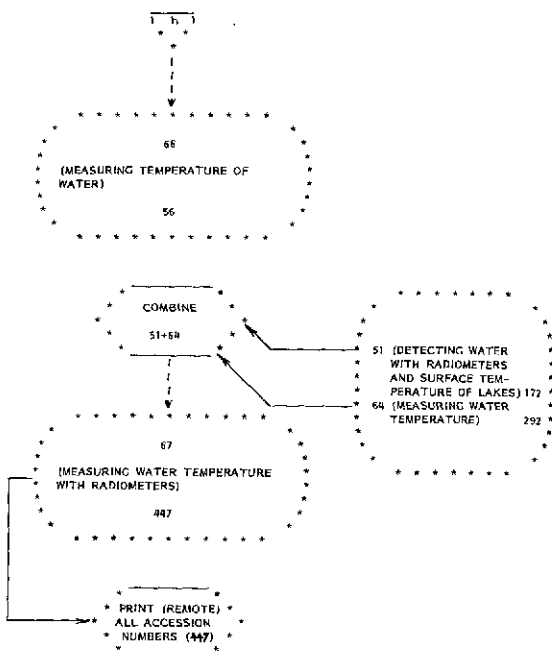
CONTINUE
DISPLAY
?

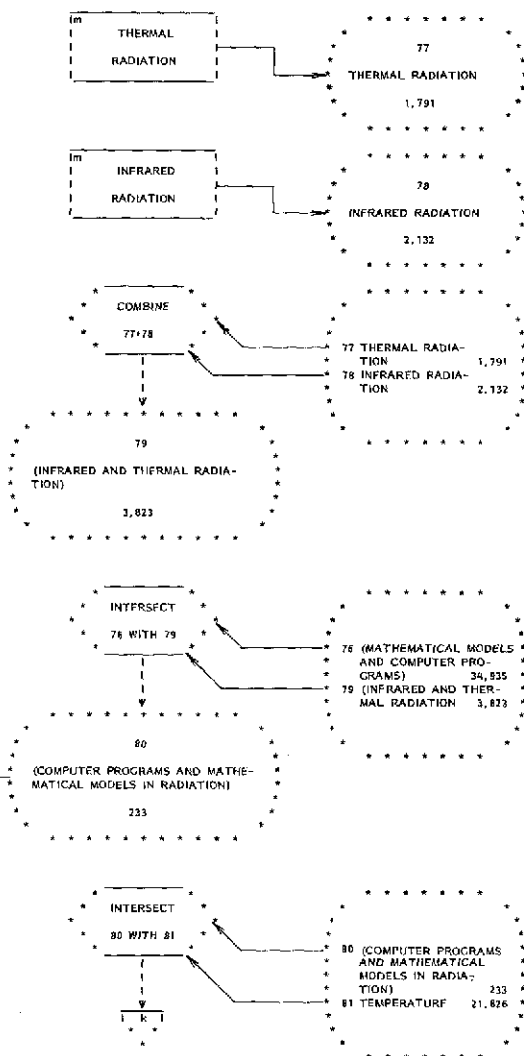
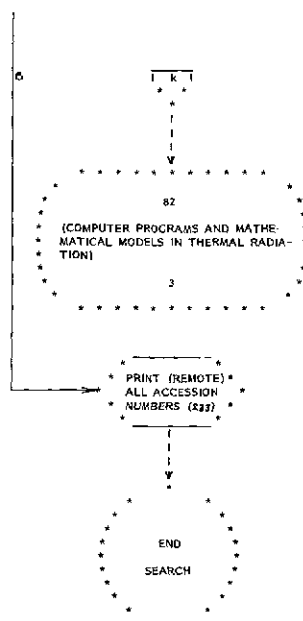
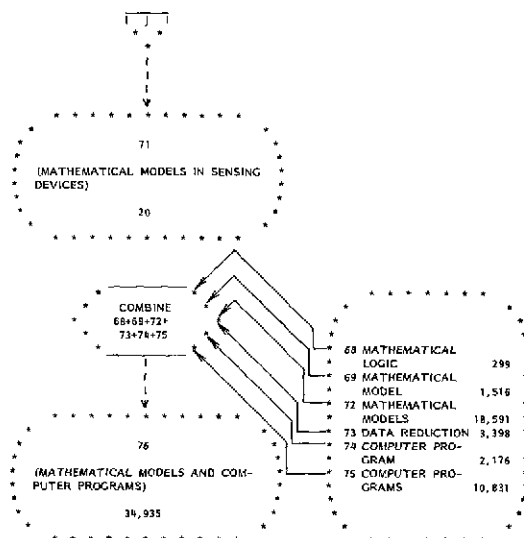
YES











LOGICAL PROCESSES EMPLOYED BY SEARCHER PJC

This searcher is an electrical engineer by training. He has had many years of industrial research, development, manufacturing, and management experience. His professional interests center on devices and systems rather than searching specific technical questions.

He began his search (Figure 6) by requesting an expansion on temperature. After reviewing all 41 terms, he constructed two sets, 1 dealing with temperature measurement and 2 dealing with temperature transducers. The first display of the expansion on "radiometer" indicated to him that the pertinent terms were just the three he used to create set 3 and he did not continue to display additional terms.

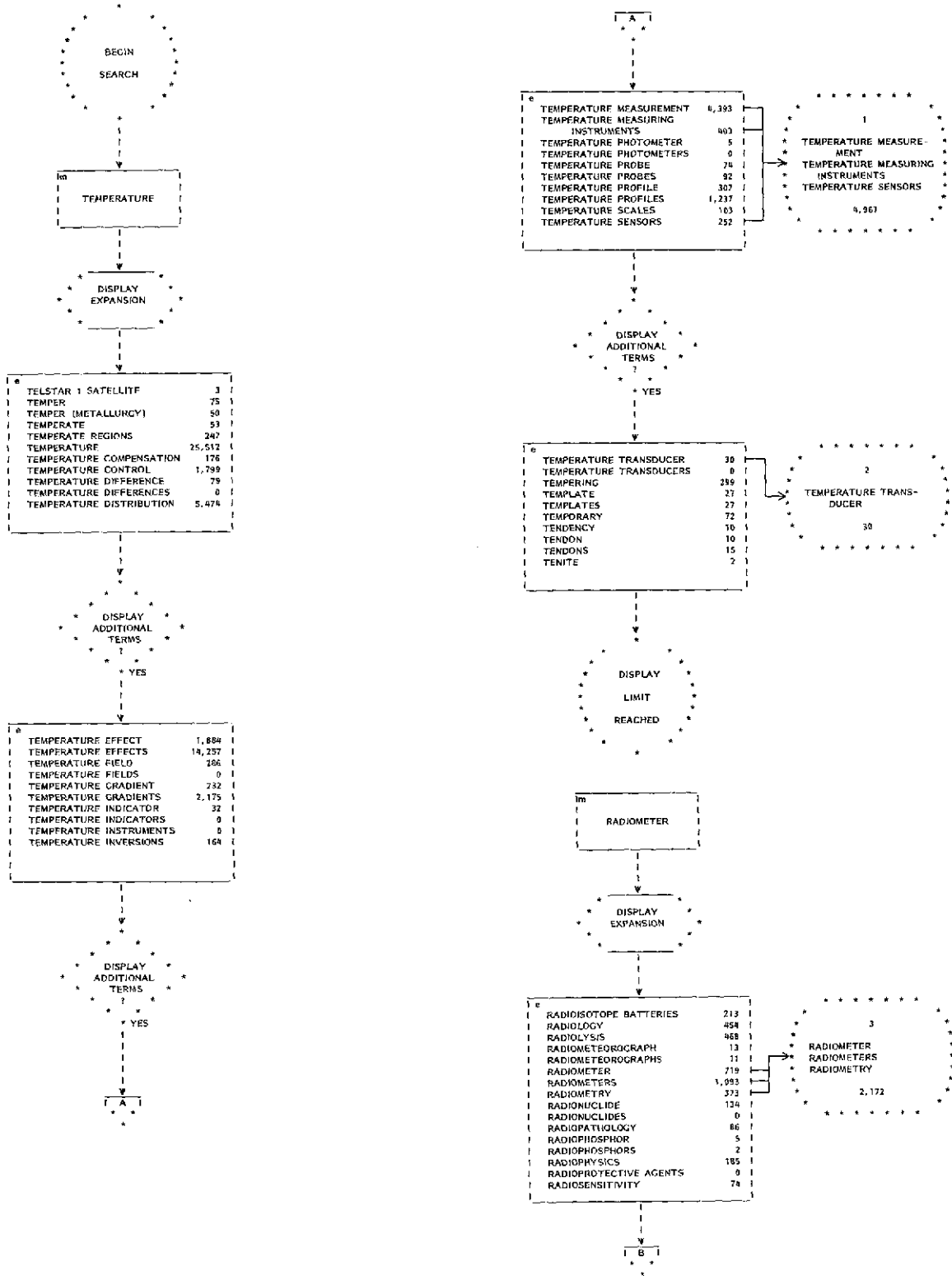
Thinking that measurement precision would certainly be a crucial aspect of the search, he selected "precision" as set 4 without first performing an expansion. He then returned to the earlier sets and performed an intersection of the union of sets 1 and 2 with set 3. The resulting set, 5, dealing with radiometers for temperature measuring, contained 147 items. Fearing that this was too small a group to represent the real content of the file, he tried a slightly different tack. He intersected a union of the first three sets with set 4. The resulting set, 6, contains documents which should deal with the precision with which one can make temperature measurements but only partially with radiometers. Only 19 citations were so identified. Examination of the full citation for two of these indicated that they were not particularly pertinent. Nevertheless, he asked that these 19 citations be printed in full and considered part of the final hit list. Moving then to the idea of the mathematical models by which the radiometer indications were interpreted, searcher PJC asked for an expansion on "mathematical model" and selected two terms from the first and final display to create set 7. He then selected set 8, "temperature measurement," and set 9, "temperature distribution," without examining an expansion. Performing an intersection of set 7 with the union of sets 8 and 9, he obtained set 10 with 375 postings dealing with mathematical models for temperature measurement.

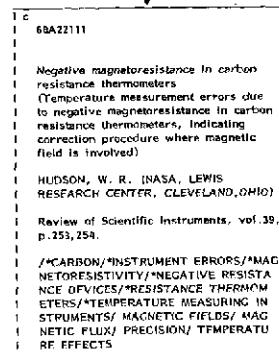
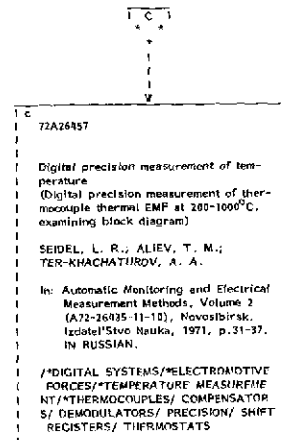
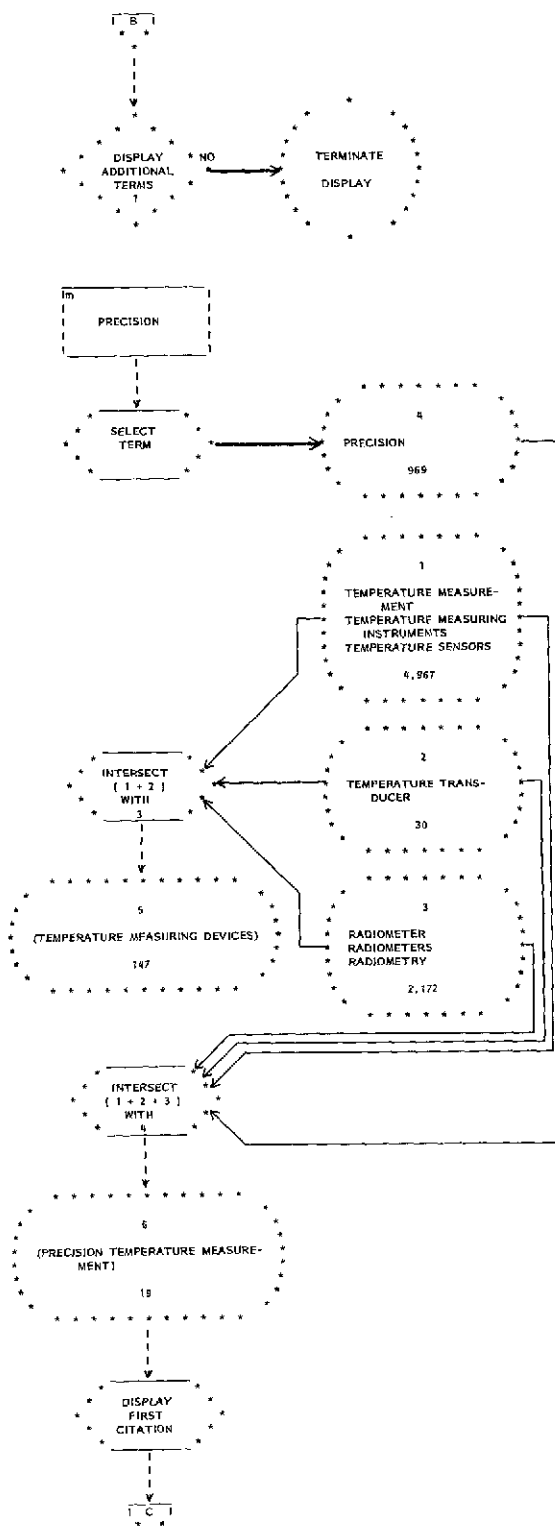
He then selected the term "data reduction" (set 11). Backing up for the moment, he tried an intersection of set 7 with set 9 to obtain set 12, a somewhat more restrictive form of set 10 with 298 postings. Documents in set 12 should deal with mathematical models for temperature distributions. Apparently searcher PJC reasoned that the restricted form contained sufficient data from which one could sift a suitable number of pertinent hits since he used set 12 as the basis of further investigation and dropped set 10 from further consideration. Before continuing in this vein, however, he intersected set 11 with set 1 to obtain 53 citations (set 13) dealing with data reduction of temperature measurements. Examination of the first three citations revealed them to be of little or moderate interest. When, however, an intersection of set 12 with set 4 gave no hits, he decided to print set 13 as part of his final hit list.

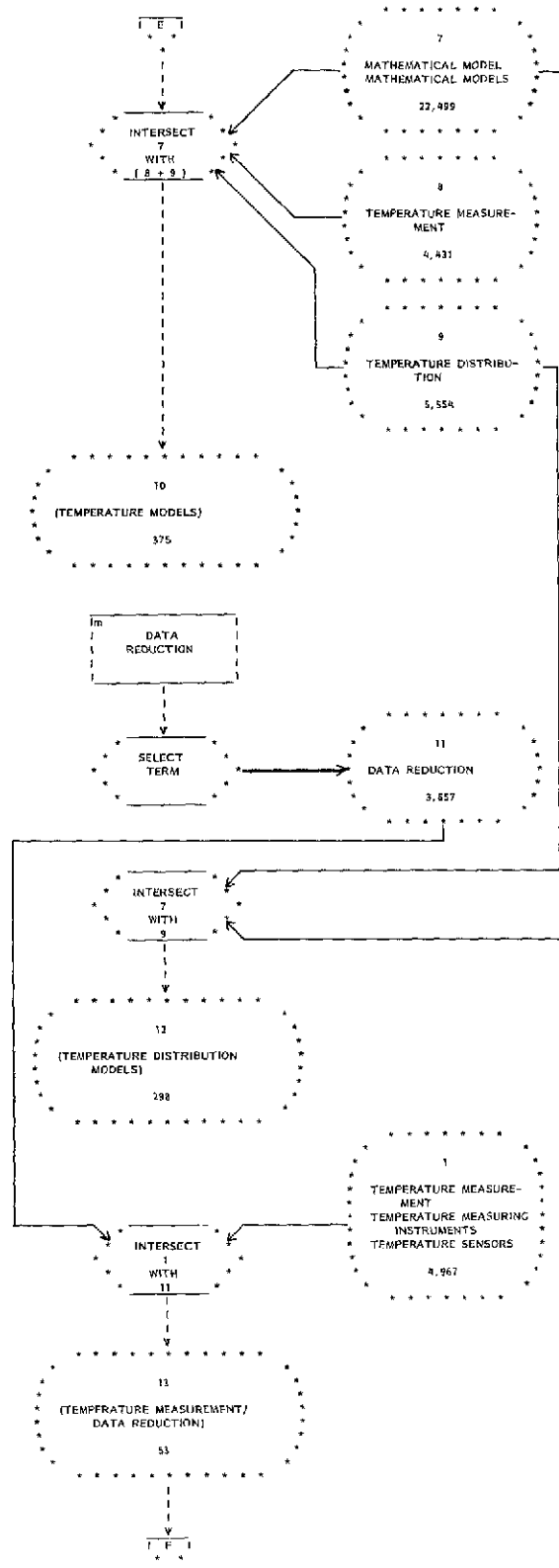
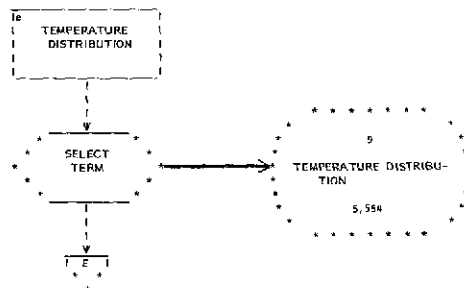
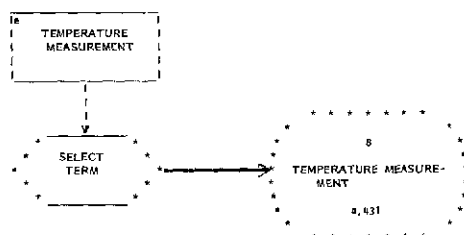
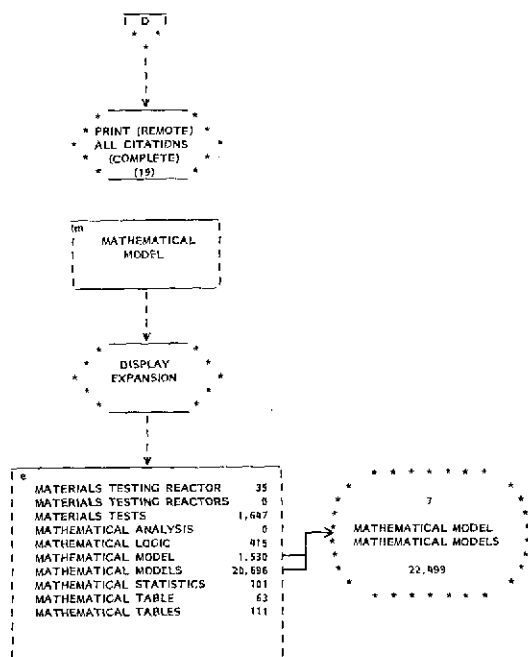
Taking one additional tack, he selected "remote sensors" and intersected this with the union of sets 1, 8, and 9, all dealing with temperature measurement. This produced set 15 with 107 hits which he chose to add to his final hit list.

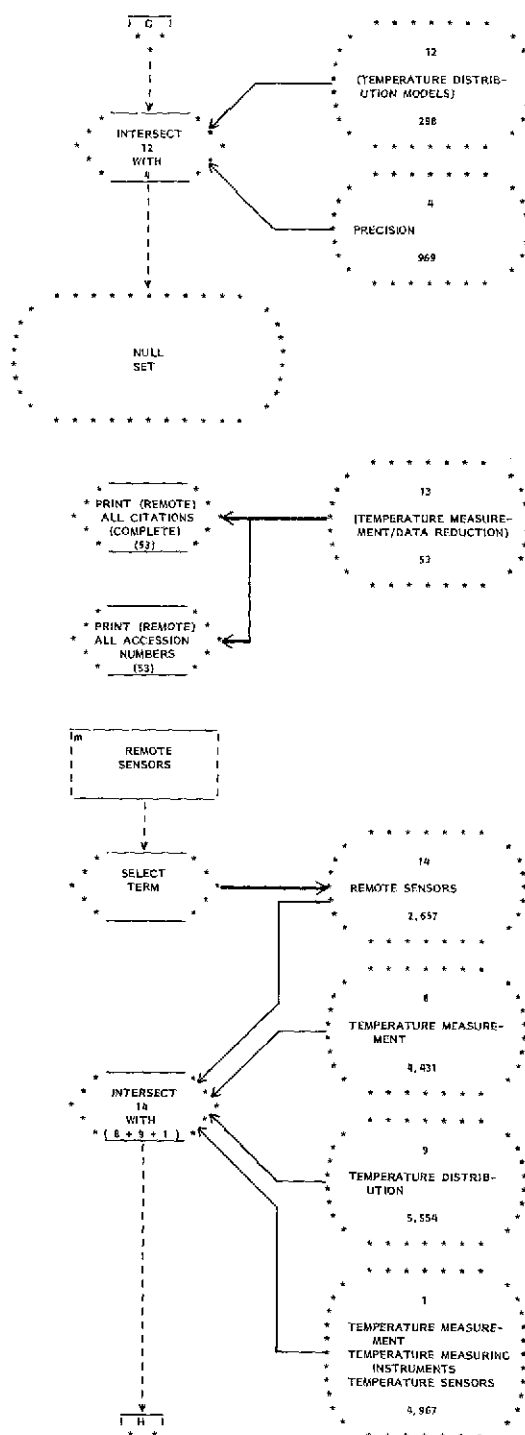
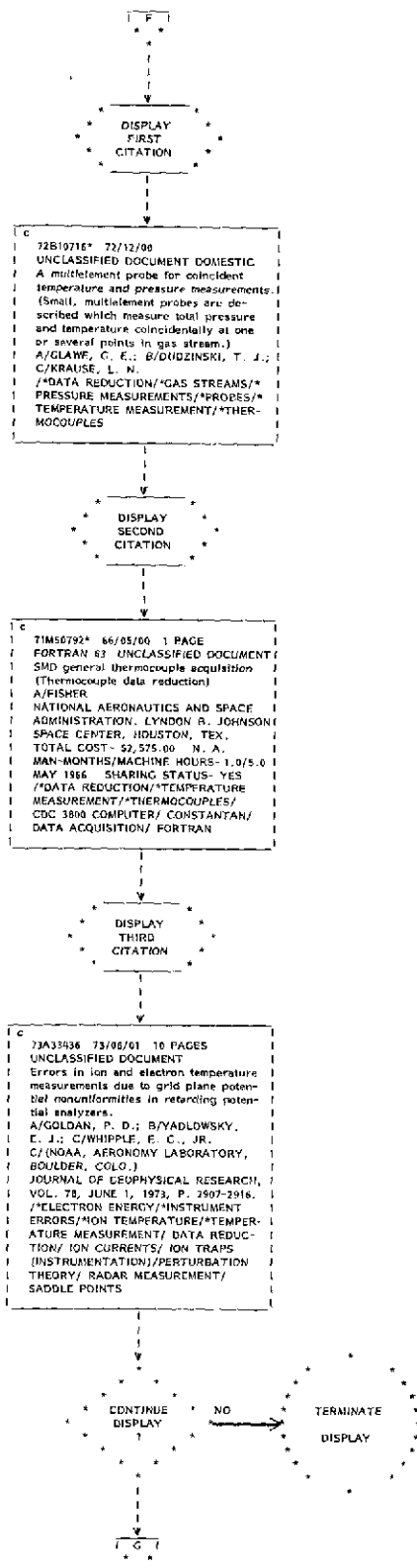
Of the 44 good hits found (out of a total of 179) by this search, only two were not found in at least one other search. Despite the limited promise of the citations he examined, about 25% of the hits in the search turned out to be relevant documents. Overall recall, however, was fairly low.

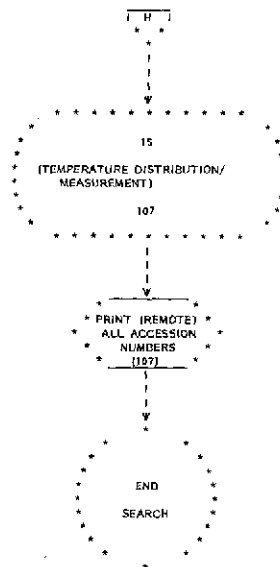
FIGURE 6











LOGICAL PROCESSES EMPLOYED BY SEARCHER SFS

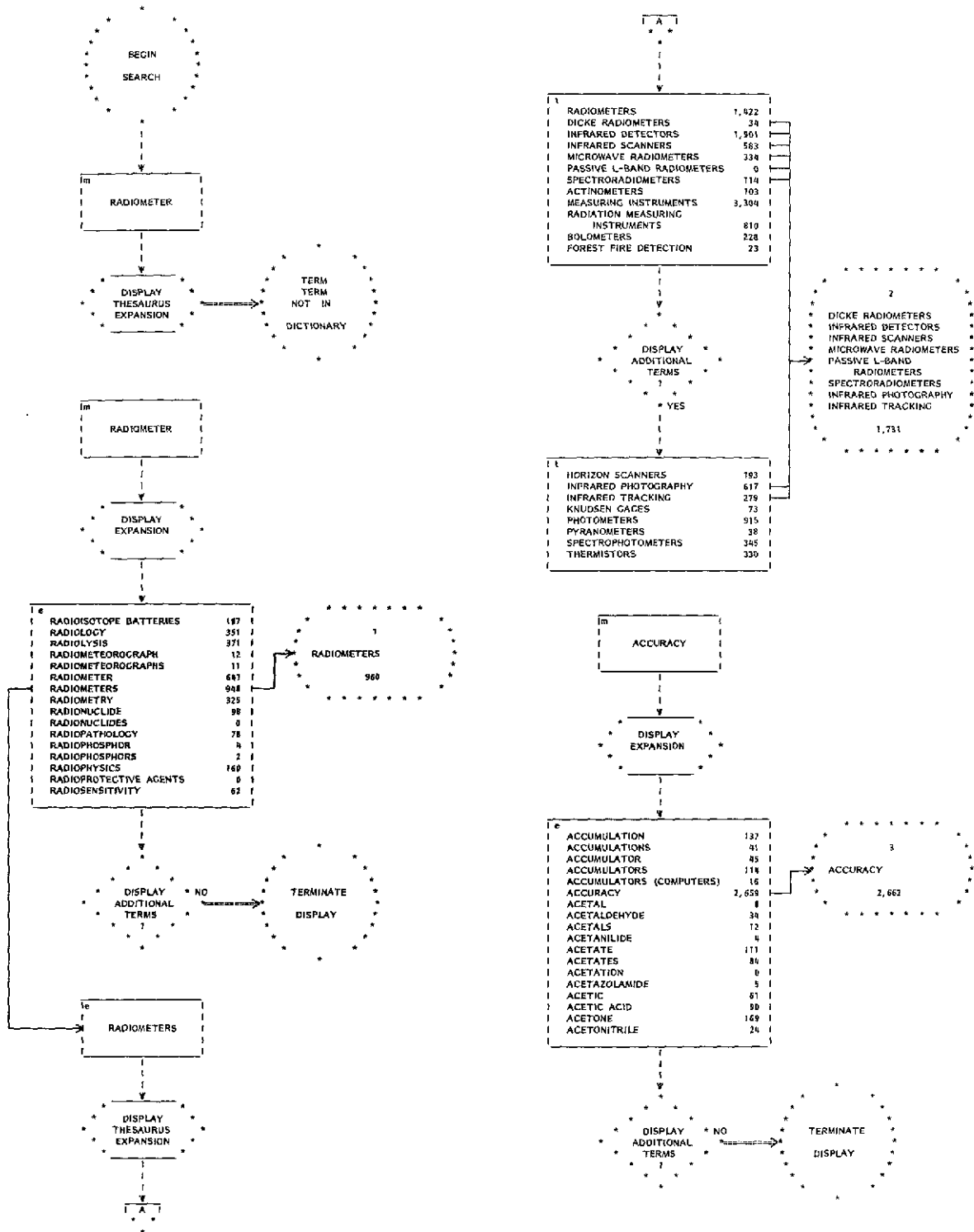
Searcher SFS is a young, graduate industrial engineer with a specialization in production but with no industrial experience. He is familiar with searching procedures, the NASA data base, and with RECON.

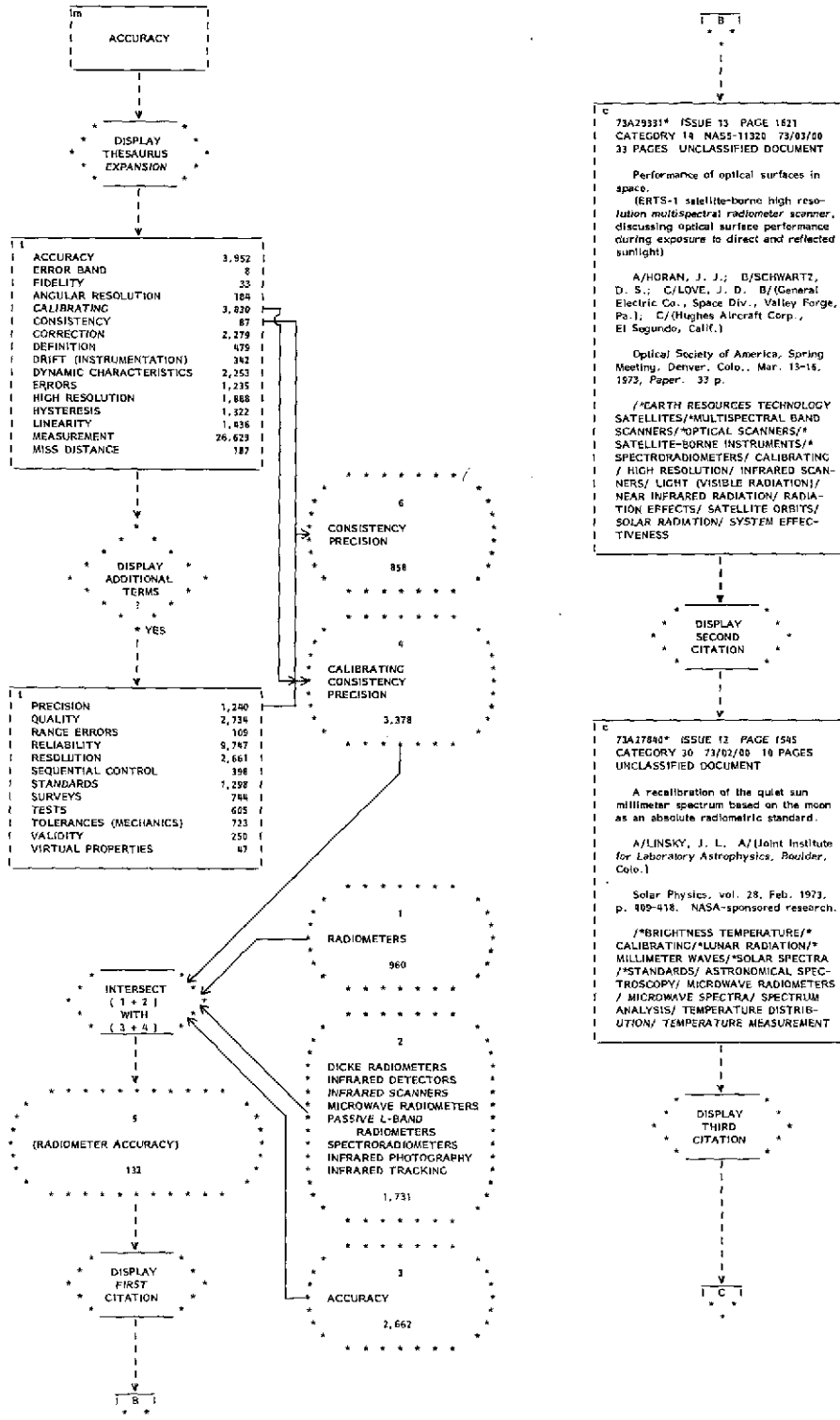
As seen in Figure 7, he began by attempting a hierarchical (thesaurus) expansion on the basic instrument used to make the measurement of lake temperature from an overflying aircraft (radiometer). Since the term did not appear in the singular in the dictionary, he asked for an expansion of that portion of the alphabetical dictionary. In this fashion he found that thesaurus expansions are generally made about plural forms. He selected radiometers as a set and then went on to perform the expansion he had originally set out to do, i.e., on radiometers. From the resulting display of 20 terms he grouped eight to create set 2.

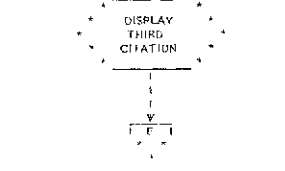
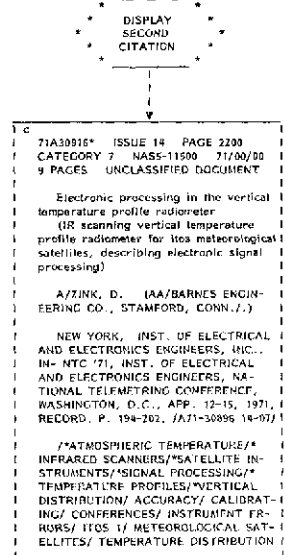
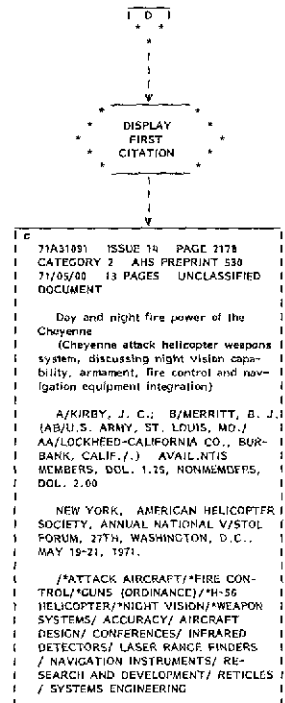
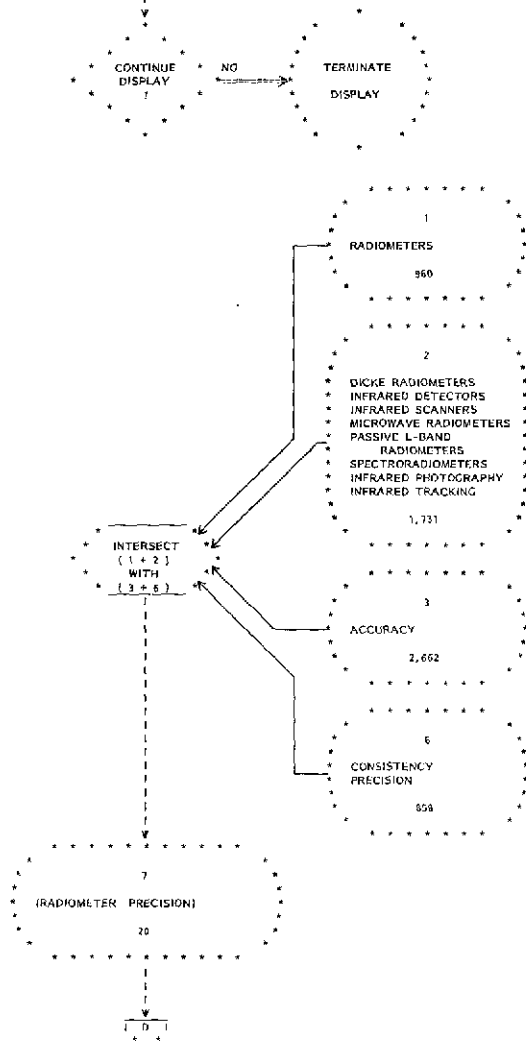
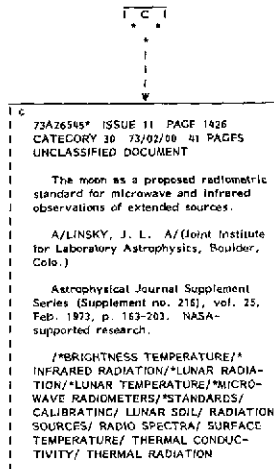
Since the customer had been concerned about possible accuracy of measurement, searcher SFS next asked for an expansion on the word "accuracy." Other than the term itself (set 3), nothing else in the display appeared to be of interest. He then tried a thesaurus expansion on "accuracy." From these displays he selected three terms from which he created set 4. He then intersected a union of set 1 and 2 (radiometers) with a union of sets 3 and 4 (accuracy). This produced 132 hits. Upon display of the first three citations, he decided that the index term "calibrating" was responsible for retrieving these citations which were not directly pertinent. He therefore dropped calibrating from set 4 thereby creating set 6. He then repeated the intersection (this time with set 6 instead of set 4) to obtain set 7 with only 20 hits. The first three citations from set 7 were only slightly more pertinent than those from set 5. This indicated to him that if he were to find a significant number of pertinent documents, a new tack would be required.

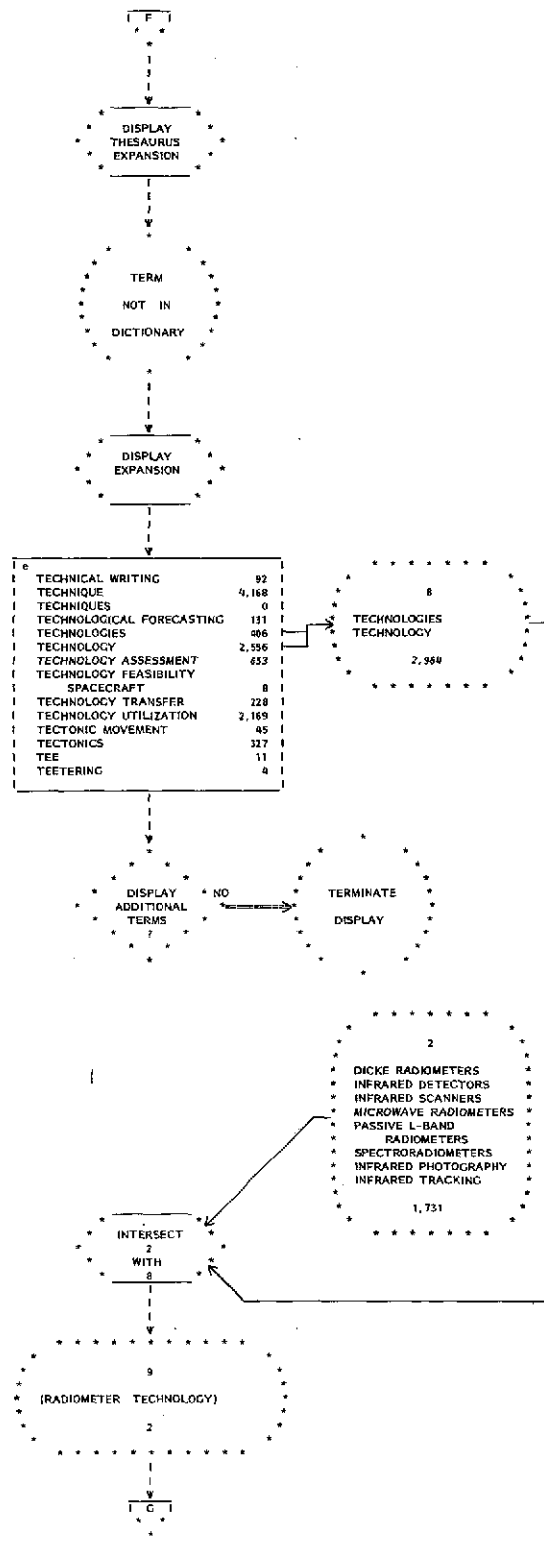
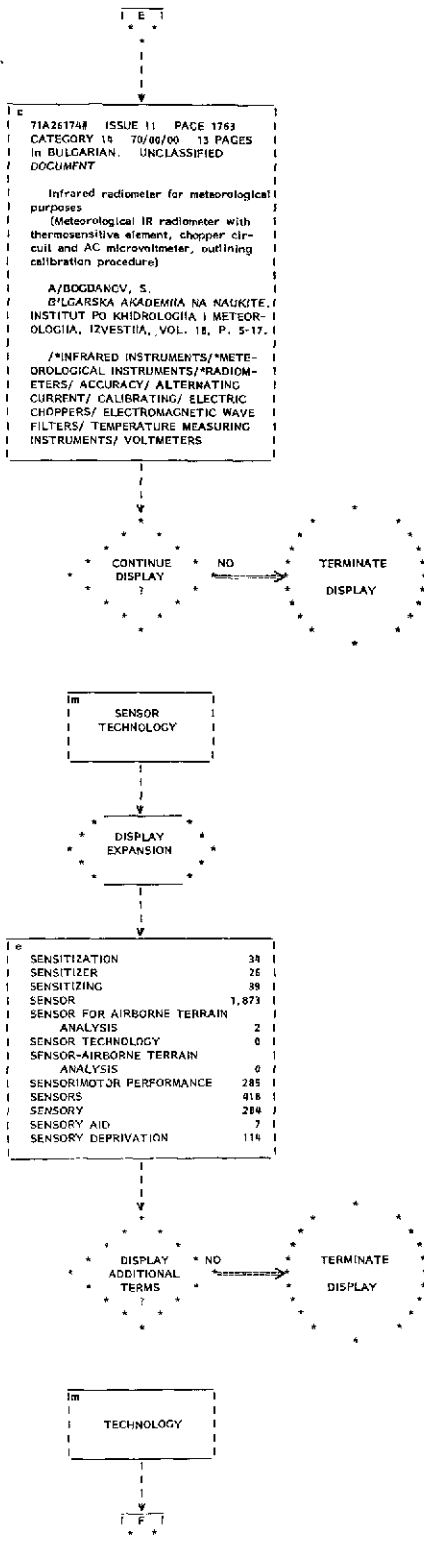
He tried an expansion on sensor technology but was disappointed with the neighboring terms and therefore discontinued the expansion. He then decided to perform a thesaurus expansion on "technology" with the result that he learned the term does not appear in the thesaurus. A dictionary expansion on the same term showed it did appear in the pre-1968 file with 2,556 postings and that "technologies" was the correct term for the post-1967 file. From these two terms, he created set 8. This he then intersected with set 2 (the thesaurus expansion on radiometers) and obtained only two hits which he displayed. They were not pertinent so he decided to accept as his final hit list the 132 items from set 5, knowing of course that the relevance of at least the first three citations was not good. He later redid the search along essentially the same lines and got a total of 160 hits. The latter items were the ones evaluated. Some 60 items from this group were regarded by the experiment director as pertinent. Thirty-one of these were not found by any other searcher. The 60 pertinent items found in this search can be compared with the total number of pertinent items found in all searches (388) to point out that while the relevance is high, the recall is only 15.5%.

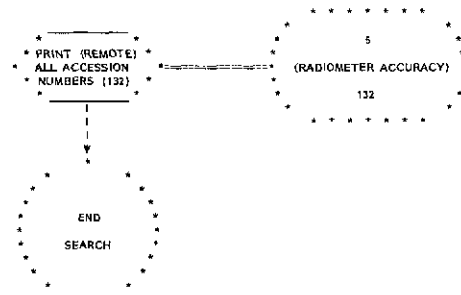
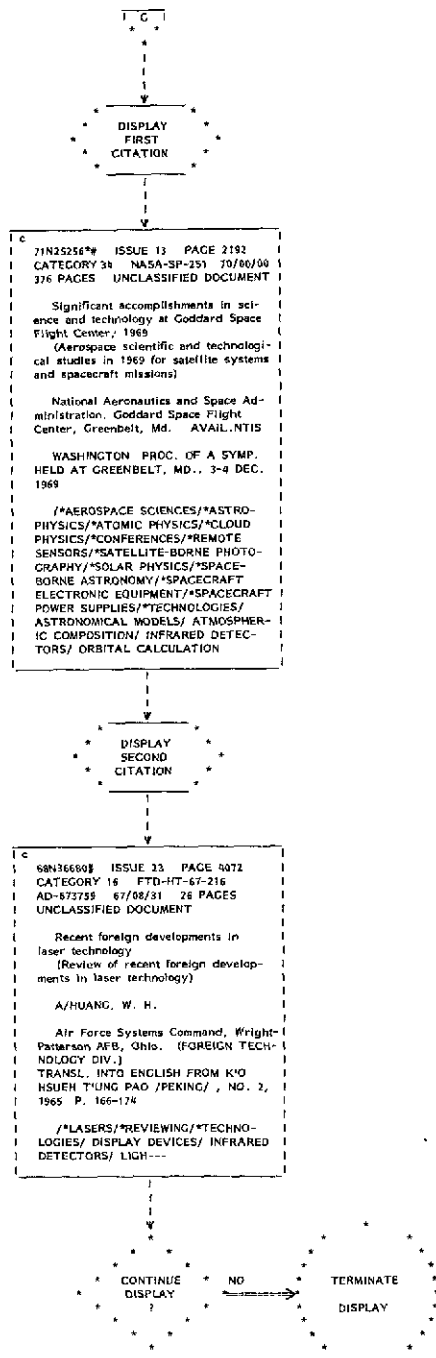
FIGURE 7











LOGICAL PROCESSES EMPLOYED BY SEARCHER CLN

This searcher is an aerospace engineer with a research and development background (primarily in structures) and a familiarity with the data base and with RECON.

He began his search (Figure 8) by requesting an expansion on "radiometer." Although he only looked at the initial display, he selected four terms centering around "radiometer" from which to create set 1. He then requested a thesaurus expansion on "radiometers" selecting 9 from which he created set 2.

His next idea was to examine the potential of terms related to thermal pollution. The dictionary expansion, he concluded, was not very productive so he tried thesaurus expansion. From this he took 13 terms to form set 3. One of the other terms in the expansion, "Lakes," seemed to him to offer an opportunity to examine a very pertinent aspect of the problem. He collected 9 terms from the thesaurus expansion to form set 4. Then, approaching the crux of the matter, he requested an expansion on "water." Set 5, taken from this expansion, deals with "water," "water intake(s)," and "water management" and has some 8,700 postings.

He then approached the idea of "remote sensing" but when the term was not found in the thesaurus dictionary, he abandoned the idea and proceeded to intersect the union of sets 1 and 2 with the union of sets 3, 4, and 5. The resulting set, 7, should contain documents dealing with radiometry of water systems. Set 7 consists of 230 items. The searcher displayed the accession numbers of these and then displayed the full citation for a new N-document and an older A-document. These appeared to be approaching the core of the matter; in particular, the first item seemed quite pertinent.

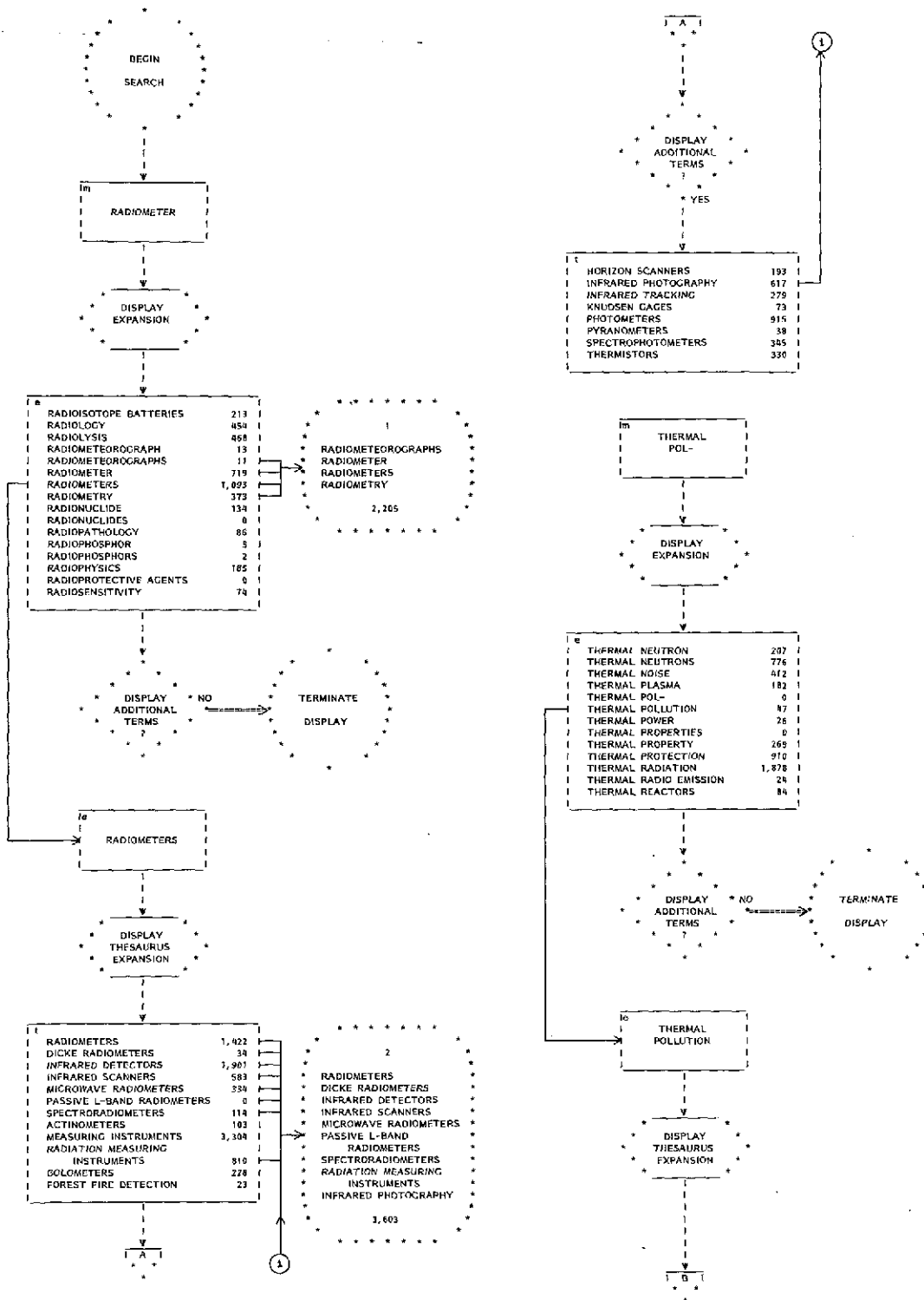
The searcher next left these documents and proceeded to examine another aspect of the problem statement; a mathematical model of the emission, transmission, and reception of radiant energy. He began by requesting an expansion on "mathematical model" and selected two terms from among the 24 that he examined. Out of these he created set 8. Since the customer, as indicated by the problem statement, was interested in solving any mathematical model deemed appropriate by computer, the searcher next asked for an expansion on "computerized." The first display convinced him that in asking for an expansion on "computerized," he had passed the point in the dictionary where the terms of most interest lay. He therefore asked for an expansion on "computer" and from this selected four terms with total postings of 16,367 as set 9.

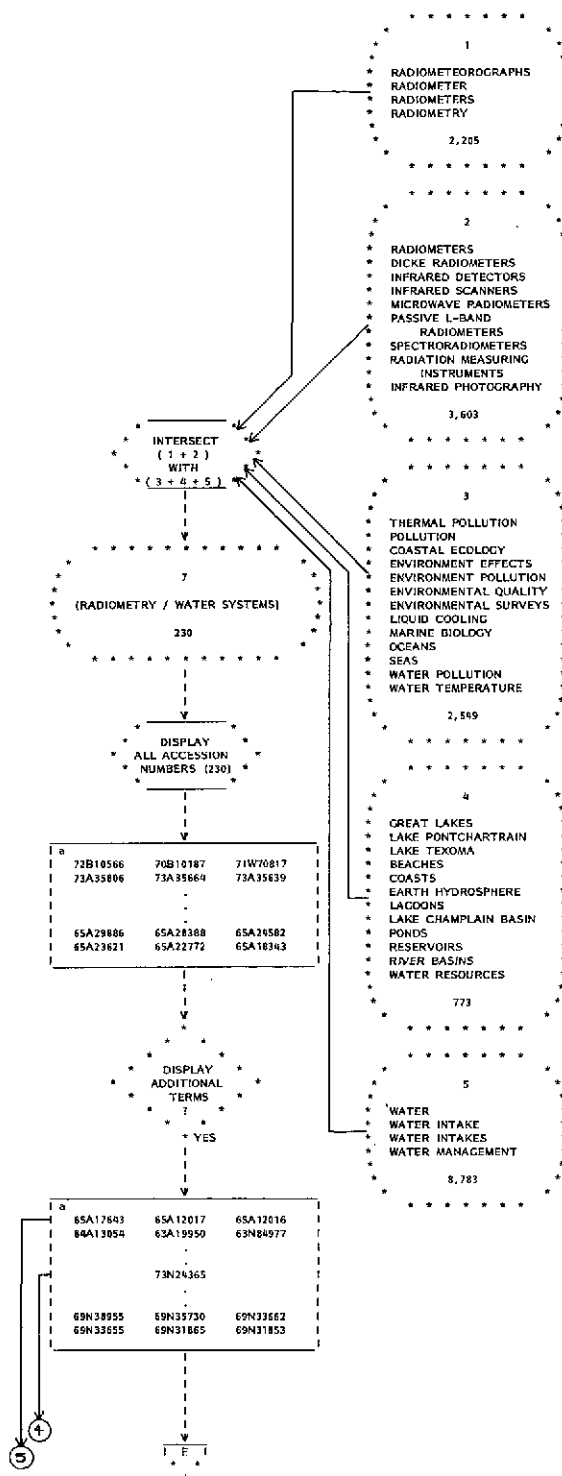
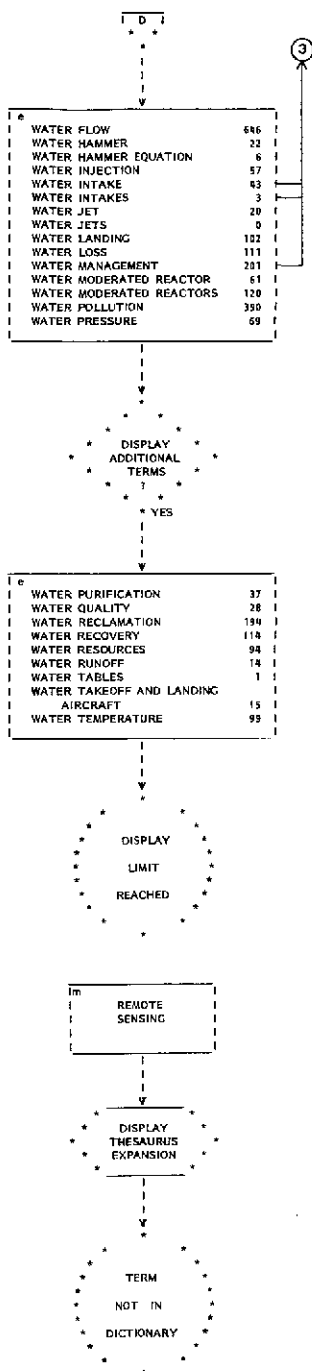
Continuing with the idea of computerized processing of data according to a given mathematical model, the searcher asked for an expansion on "data." He formed set 10 with 12,745 postings from 5 of the terms in this expansion which, incidently, he carried through to 40 terms. Next, he asked for a thesaurus expansion on "Data reduction." The results he did not care for. He then intersected a union of sets 8, 9, and 10 with a union of sets 1 and 2 to obtain set 11 of 288 postings which dealt with modeling and processing of radiometer data. Apparently, the number of documents in this set was somewhat larger than he reasoned might be really pertinent so he performed

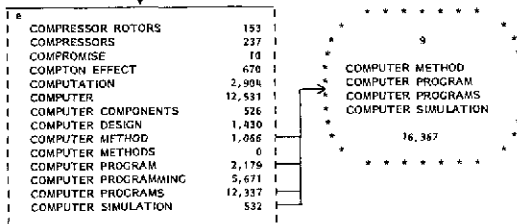
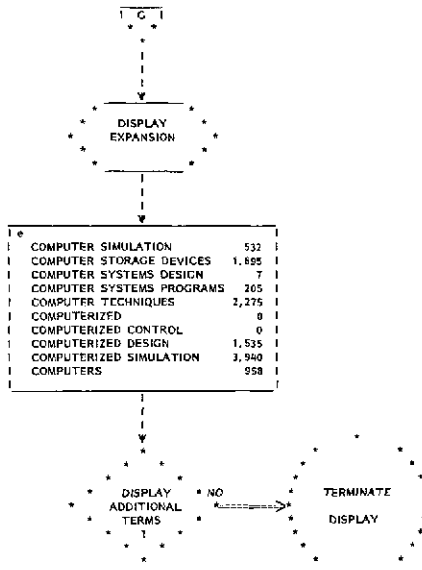
$(1+2) * (8+10)$ and got set 12 with 256 hits; $(1+2) * 10$ and got 202; $(1+2) * (8+9)$ and got 102. Each of these subsets of 11 were fairly large, indicating substantial overlap in indexing. He tried to get somewhat more specific hits by performing $(1+2) * (8+9) * 10$ which yielded set 15 with 16 hits.

Backtracking, the searcher then combined sets 7 and 11 to obtain set 16 with 504 documents dealing generally with radiometry, water systems and data analysis. At this point, he decided he would go back to the thesaurus expansion of "radiometers." From this expansion, he selected five terms, instead of the nine he used to form set 2, to create set 17. He then performed an operation similar to the creation of set 11, except that he replaced set 2 by set 17. This gave him 164 hits (compared with 288 in set 11). He added to this set, set 18, the 230 documents from set 7 dealing with radiometry and water systems to form set 19 with 387 documents. This he called his final hit list. Ninety-seven of these hits were considered good (relevant) by the project director. Eighteen of these were not found by any other searcher.

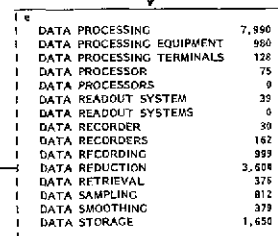
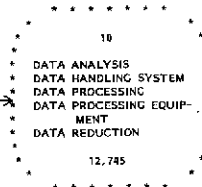
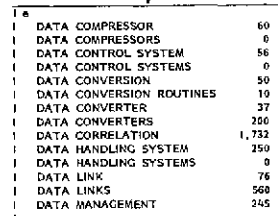
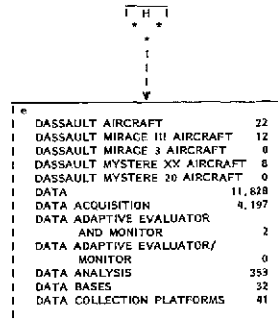
FIGURE 8



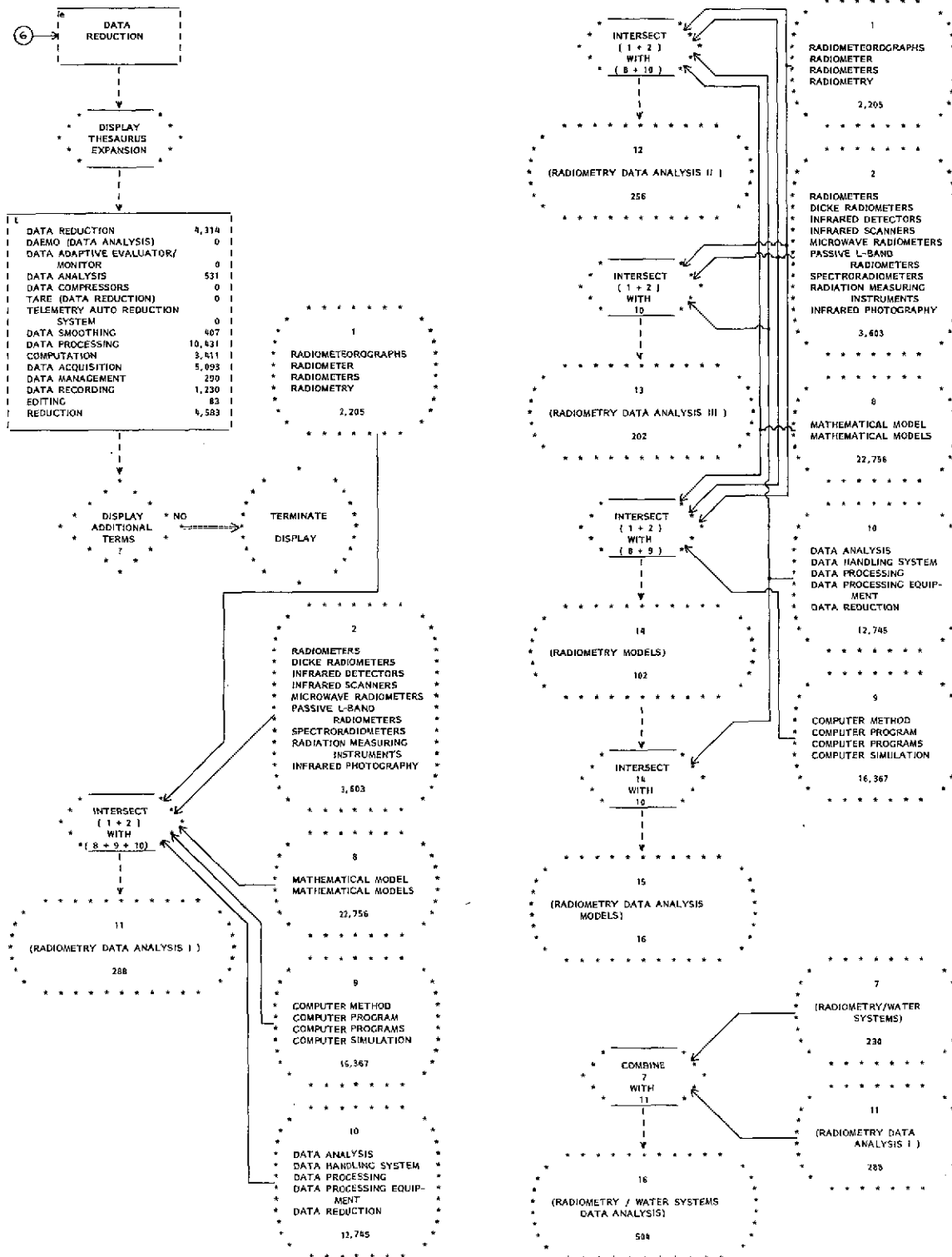


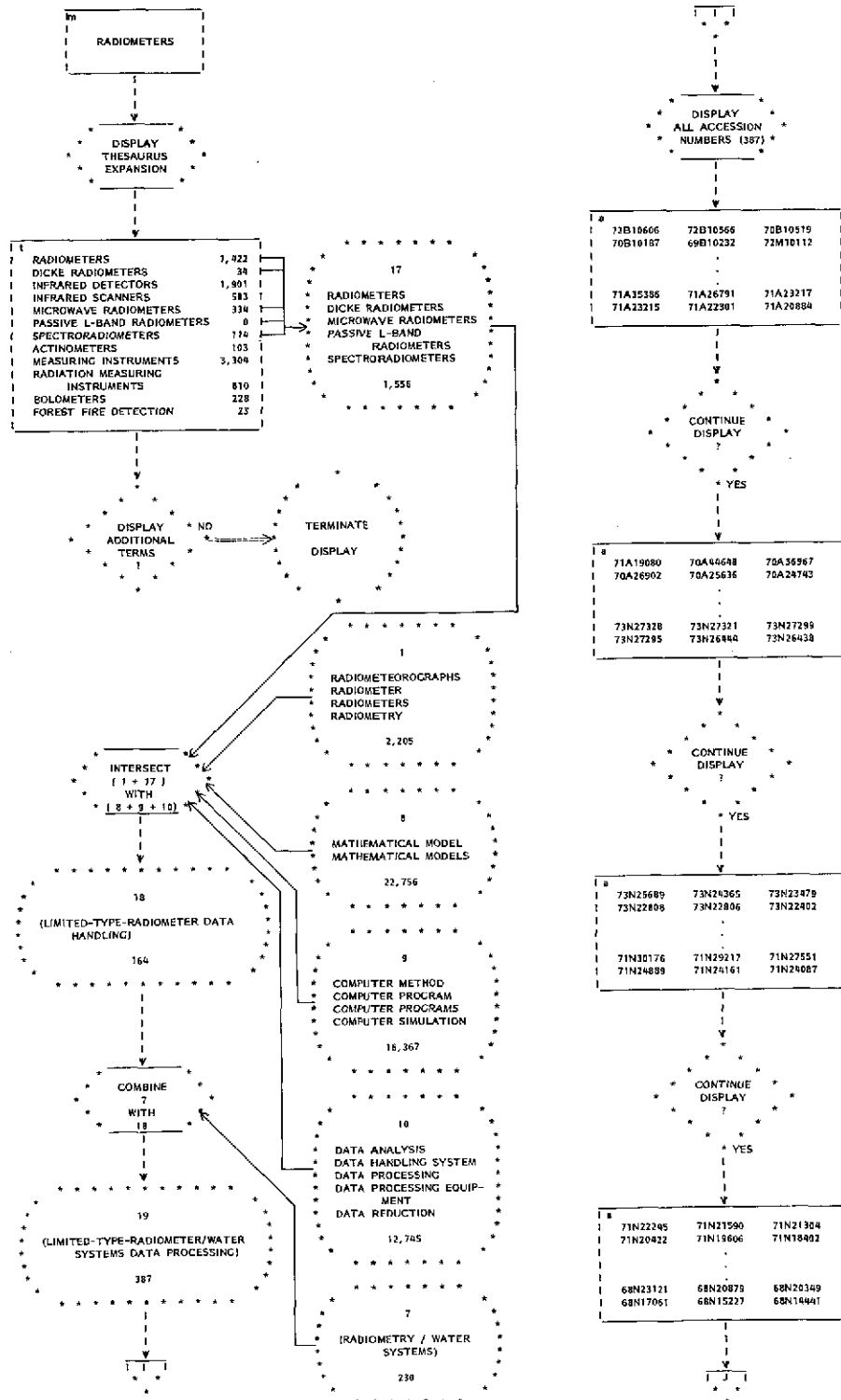


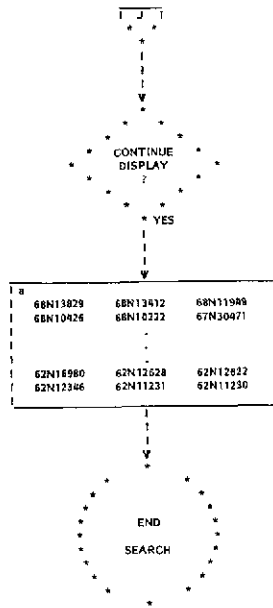
1 H 1



6







CORRELATION OF SEARCH RESULTS WITH SEARCHER'S BACKGROUND

Results produced by the eight significant searches on this subject can be divided roughly into two groups: those which gave at least 87 good hits (average 101) and those with 60 or fewer good hits (average 48). The first group of searches was written either by experienced searchers or by engineers with a working knowledge of the subject area. Searcher AWL compensated for his lack of familiarity with the subject area by doing a rather broad search. His recall was fairly good (22.4% compared with 31.2% for the best search in the group) but his relevance was low (13% compared with 36.6% for the best in the group). Searchers TRP and CLN, the most experienced in the group on the NASA file, both have backgrounds in areas related to the subject area but not directly on it. Their searches gave good recall and relevance and, in addition, employed considerably simpler logics than the batch search or the one done by WWS. The latter two searches were done by working engineers with backgrounds fairly close to the subject area. From their results it seems that being more diligent usually will lead to a slight improvement in recall percentage.

The other four searches were performed either by searchers with no relevant education or experience or by searchers with little regular search experience. Searcher SFS, closest of the four to the subject area by background and search experience, also did the best job in terms of relevance and recall. While his search had the best relevance of any search done on this subject, his unfamiliarity with the subject undoubtedly contributed to his failure to follow synonymous paths to the other documents on the subject and thus limited his recall.

CORRELATION OF SEARCH LOGIC WITH RECALL ACHIEVED

As the initial investigation into the effects of search logic on the recall achieved, the reasons searcher WWS failed to retrieve some of the documents found in the batch search were studied. It was found that the excluded documents were generally indexed under one of three groups of terms:

- (1) INFRARED INSTRUMENTS
INFRARED RADIATION
INFRARED IMAGERY
- (2) THERMAL POLLUTION
- (3) SURFACE TEMPERATURE

It was then found that the terms in group (1) do not appear in the thesaurus expansion of radiometers. Thus, unless the searcher happened to think along these lines, the terms would not have been suggested to him otherwise. A similar situation exists with respect to group (3). This term does not appear under a thesaurus expansion of "temperature." Failure to include "thermal pollution" seems to have been just a mental lapse on the part of searcher WWS.

No comparisons with the batch search are possible for pre-1968 documents because the batch search excluded these. Thirty-six good hits were found by the WWS "A" search for this period.

The WWS "A" search found 58 good hits not found by the 152-term batch search; only 27 good hits were the same in both searches. The reason, apparently, is that the batch search employed a double intersection along the following lines: (temperature measurement) * (radiometers) * (accuracy). Examination of 24 of the good documents found by WWS but missed in the batch search indicated that these documents were not indexed under a term synonymous with accuracy such as calibrating, correction, digital filtering, etc., all of which were used in the third group of the intersection. They were, therefore, not selected.

In general one must conclude that, for this topic at least, high recall can only be obtained by:

1. employing every conceivable term related to the subject
2. single intersections only
3. extensive manual sort down.

One may also conclude that while each of the searches employed a valid search logic, those that employed the most terms and the most different approaches to the same topic were the most successful in terms of recall. Some approaches of course favored higher relevance than others. Searches employing these approaches only appeared to be highly (>30%) relevant but were so at the expense of recall.

RELATIVE EFFECTIVENESS OF INTERACTIVE AND BATCH SEARCH MODES

A second experiment, this one attempting to access the relative effectiveness of the interactive and batch search modes, was also conducted. Engineering time to conduct the search was compared with search results as measured by good hits. As shown in Table 2, engineer CLN got essentially the same results (using substantially the same logic) by RECON as he did by the batch process. The search time was almost identical. On another question it took him somewhat longer to search via RECON (probably because of long question response times). Again the good hits were substantially the same items.

On question #4, PJC took 25% more time to do the problem on RECON as by the batch process, while searcher SFS took only half as long to get a few more good hits on question #5. In the latter case, he spent a good deal of time working out the strategy prior to the batch search. When he went to the RECON terminal, he knew generally how he was going to attack the problem.

Generally, the effectiveness of a RECON search was not better than that of a batch search conducted by the same individual using the same amount of engineering time to prepare the search. One can also correlate reasonably well, using a logarithmic function, the engineering time spent in preparing the search with results (recall) achieved.

One final piece of data bearing on this question is the study made of searcher TRP's batch and RECON searches during a given period (Tables 3 and 4). Some 15 batch searches and 10 RECON searches were studied. It was found that his RECON searches achieved about twice the average relevance of his batch searches at the cost of about 35% more engineering time per search.

TABLE 2

Search Question: *Airborne Thermal Mapping of Water Surfaces*

	ENGINEERING TIME	COMPUTER TIME	GOOD HITS
Manual - Batch FOS	108 min.		104
RECON: TRP	163 min.		97
MRN	180 min. total 51.91 at RECON		39
WWS "A"	170 min.		121
WWS "B"	25.43 min.		8
PJC	64 min		44
SRW	102.51 min.		48
SFS	60.8 min.		60
CLN	124.35 min.		97
AWL	184 min.		87
Search Question A (CLN)			
Batch	90 min.		*
RECON	91.82 min.		*
Search Question B (CLN)			
Batch	75 min.	1.13 min.	101
RECON	105 min.		*
Search Question C (PJC)			
Batch	60 min.	3.85 min.	689
RECON	75.51 min.		*
Search Question D (SFS)			
Batch	90 min.	1.45 min.	158*
RECON	42.72 min.		222

*

Output substantially the same - By substantially the same we mean about the same hits as well as the same number of hits.

TABLE 3
BATCH SYSTEM

Statistics Based on TRP's NASA Searches

	<u>Total Engineer Time (prep. & eval.) in hrs.</u>	<u>Total Hits</u>	<u>Evaluated Hits</u>	<u>% Relevant Hits</u>
1)	2.5	256	44	17.2
2)	3.5	409	31	7.6
3)	2.5	257	64	24.9
4)	3.0	112	5	4.5
5)	3.5	385	63	16.4
6)	3.5	207	42	20.3
7)	2.5	71	64	90.1
8)	2.5	167	120	71.9
9)	3.0	275	77	28.0
10)	3.0	216	35	16.2
11)	3.0	447	54	12.1
12)	3.0	353	49	12.9
13)	3.5	74	7	9.5
14)	3.0	169	60	35.5
15)	1.5	143	15	10.5
Average				
	= 2.9	= 236.1	= 45.6	= 25.2%

TABLE 4
RECON

Statistics Based on TRP's NASA Searches

	<u>Total Engineer Time (prep. & eval.) in hrs.</u>	<u>Total Hits</u>	<u>Evaluated Hits</u>	<u>% Relevant Hits</u>
1)	4.0	219	100	45.7
2)	3.0	236	108	45.8
3)	3.0	183	101	55.2
4)	3.0	457	32	7.0
5)	5.0	no data	53	no data
6)	4.0	111	51	45.9
7)	5.5	no data	150	no data
8)	3.5	166	38	22.9
9)	5.0	no data	183	no data
10)	4.5	483	319	66.0
Average				
	= 4.1	= 265	= 113.5	= 45.6%

RELATIVE COSTS OF BATCH AND INTERACTIVE SEARCH MODES

Cost Structure for the Batch Retrieval System - STRC-IVS

Direct Costs

Engineering Time

This is the time spent by an engineer (or search analyst) in preparing a search strategy for input to the computer and in evaluating the search output once his strategy has been run. Typically, this time is spent in several ways: discussing the search problem with a client, looking through the appropriate thesaurus for terms, checking those terms against a computer-produced list of authorized terms and postings, composing the search strategy, and then finally reading or scanning the hit abstracts for relevant documents. STRC cost accounting reports show that, on the average, an engineer spends 1 1/2 - 2 hours in preparing a search and 2 - 2 1/2 hours in evaluating the search output at a total cost of roughly \$35 to \$40.

Clerical Time

Clerical time for a search includes such functions as interpreting search requests, verifying search terms and postings, key-punching or keyboarding via terminal the actual search request, checking search output, retrieving and Xeroxing hit abstracts, and binding search bibliographies. Accounting statistics show that on the average, the clerical time (two hours) involved in a search costs STRC \$5 to \$6.

Xeroxing Cost

The average cost for Xeroxing of search output is roughly \$7.

Computer Cost

STRC searches run on an overnight, no-priority basis on the Triangle Universities Computation Center's (TUCC) IBM 370/165. The TUCC charges for zero priority service include the following:

- 1) CPU time @ \$350/hour
- 2) EXCP (I/O) time @ \$262.50/hour
- 3) Core use @ \$.70/kilo-byte hour
- 4) Unit record I/O transmission (i.e., printing time)
@ \$.00085/UR (line)

Accounting records show that on the average, the computer cost for a NASA search ranges from \$10 to \$20.

The total direct cost for an STRC batch search then averages \$60 - \$70.

Overhead for the NASA File

A number of overhead items are incurred in maintaining the NASA file and the costs involved are absorbed by STRC.

Monthly File Maintenance

- 1) \$50 to \$75 computer cost
- 2) \$5 clerical cost

Direct Access Storage

- 1) Rental of five 2314 disk packs @ \$9.50/pack/month

Magnetic Tapes (Back-up and Maintenance)

- 1) Approximately 50 tapes at one-time purchase price of \$9.50/tape.

Discussion of Possible Cost Structures for RECON Interactive Searches

Since STRC is not aware of any cost accounting scheme that might exist for RECON searches, one can only speculate on what the elements of such a scheme might be. It is logical to conclude that one or more of the following items might be included in the cost/search for RECON.

Computer Cost

Costs for interactive services vary from machine to machine, but it may help to give an example of the TUCC charges for interactive service on the 370/165:

- 1) CPU time @ \$1000/hour
- 2) I/O channel program time @ \$750/hour
- 3) I/O wait time @ \$375/hour
- 4) Port connect time @ \$2/hour

These costs are definitely high estimates for RECON since it runs on a less expensive machine, the IBM 360/50. The RECON charges would probably run less than 1/5 the TUCC charges and maybe as low as 1/10 for CPU time.

Communications Cost

This would include terminals, communications controllers, leased lines or dial-up ports, and modems. One example of cost in this area is that of the leased communications line to College Park @ \$150 - \$200/month. This cost would be reduced if STRC were to install dial-up RECON service @ roughly \$.28/minute. The Bunker Ramo terminal and controller now in use with RECON leases for \$345/month.

On-Line Storage Cost

It is estimated that a file the size of the NASA RECON file(s) could cost as little as \$800/month for on-line direct-access storage depending on the devices used.

People Cost

Systems and Programming Support

This includes people time at NASA and STRC in monitoring the system.

Engineering or Search Analyst Time

Here the engineering functions would be essentially the same as in batch but the amount of time involved might be quite different. The engineering functions of preparing a search would be performed at the terminal, and the time involved would depend quite heavily on the response-time of the terminal and on the amount of browsing that the engineer might engage in. Average times for this activity on experimental searches are two to three hours. In addition, the engineer might carry out some of his search evaluation at the terminal and thus reduce the amount that he normally would have to perform at his desk later. Cost for engineering time would be at least equal to the \$35-\$40 cost for batch engineering time and may even exceed it.

Clerical Cost

Obviously, with RECON, the clerical functions of preparing a search for entry into the computer would be eliminated. The only clerical time involved then would be the time spent in pulling and Xeroxing abstracts and in binding search results. This averages approximately \$3-\$4/search.

A more useful rule of thumb for the possible pricing of RECON searches would be to use a charge structure similar to that of one of the organizations (such as Lockheed or Systems Development Corporation) that offer nationwide interactive on-line search service on a number of files. This approximate charge would substitute for all of the above cost elements with the exception of the people cost. These organizations normally charge the on-line user a fixed hourly rate for his terminal connect time (i.e., the total elapsed time in the system from sign-on to sign-off). These hourly rates range from \$25-\$80 per file searched. For a file the size and complexity of the NASA file, the rate probably would be in the neighborhood of \$50/hour for file access. In addition, the user must pay the telephone company for direct-dial service (or use of the TYMSHARE network) to the computer. From STRC to College Park, the phone cost would average \$15/hour. Therefore, our hourly rate (based on a Lockheed-type pricing scheme) for RECON service might be \$65/hour.

Batch vs. RECON Cost Comparison Based on Estimated RECON Cost

If one multiplies the Lockheed-type hourly rate by the two to three average hours per search that were spent at the RECON terminal during experimental searches (see TABLE 2), then he comes out with a computer cost figure (\$130-\$195) roughly 10 times that of the STRC average batch cost (\$10-\$20). Since experimental results show that the engineer does not spend less time in preparing a RECON search and may likely spend more, the cost for his time will be at least what it is for a batch search (\$35-\$40). Although there may be some savings in the clerical cost, they are miniscule (\$2-\$3) in comparison with computer and engineering costs and Xeroxing costs should remain the same. The total estimated cost for a RECON search (\$175-\$245) is then roughly two or three times that of a batch STRC search (\$60-\$70).

It also should be mentioned that the Lockheed-type hourly rate that was used in estimating the RECON costs may be unrealistically low as it is based on a rather high volume of usage and it may exclude some very high overhead costs.

CONCLUSIONS

Although this discussion has been attempting to compare apples and oranges and the author is fully aware of that, the evidence (based on similar on-line systems) at this time seems to show that the batch approach to searching NASA is two to three times more economical than the RECON approach. However, no attempt has been made in this section to judge the relative values of the results of each approach, and that has to be an important consideration in the total picture.

SUBJECTIVE USER EXPERIENCE

[The thoughts presented below were compiled by NC/STRC's most experienced RECON user. While they are those of a single individual, they do seem to parallel rather closely the experiences of other RECON users on the staff at comparable points in their acquaintance with RECON.]

One of the first problems that faces a novice at the RECON terminal (if a non-typist) is that of the keyboard. Most of the difficulty disappears after a few hours of practice, but the problem of where the "F" (or some other elusive letter) went remains to plague the user for many months. This produces a tendency to use EXPAND functions for no other purpose than to select a term which appears to be too much trouble to type. I suspect that even an accomplished typist might be tempted to do the same, particularly when the desired term is an especially long one or when the exact spelling (singular or plural, hyphen or space, etc.) is not certain.

The big problem, which although purely psychological is not unimportant, is the nagging suspicion that the fumbling approach to initial use will result in the grinding halt of a multi-million dollar computer probably accompanied by sparks and smoke at the other end of the phone line. Initial instruction to the operator should make clear that nothing could be done either to harm the system or even to produce a ton or two of computer printout. The worst that can happen is that the operator is informed, "IFOOL, WRONG..."

The uncertainty of what will happen under given circumstances leads the novice to make much more use of the EXPLAIN function than is called for. Although much of the material under command of the EXPLAIN key is basic to operation of RECON, it appears to be on one hand overly simplified and on the other hand too sophisticated for other than academic interest. Basic information concerning key functions appears to be dwelt upon with excruciating detail while practical material of use to the operator is difficult to find. As examples, the EXPAND and COMBINE functions are pursued in great detail. These functions are normally well understood before one approaches the keyboard. The error codes which are normally disregarded with a "try-again-maybe-it'll-work-this-time" attitude take up a large part of the EXPLAIN file. In practice, I doubt that anyone ever looks at error codes other than as a matter of curiosity. On the other hand, how does one determine the techniques for obtaining material from the "Corporate Source" file? Or how does one know if selections made with a range function are inclusive or exclusive of the cut-off indicated by the printer? Academic information available from the EXPLAIN function appears to completely swamp or obscure information that is of primary importance to the user.

Much of the above criticism could be remedied by the availability of a detailed user's manual which explains the operation of RECON from the beginner's standpoint. Although an in-house manual was prepared for the use of novice operators at this location, I feel that a professionally prepared operator's manual is a dire need. Many hours of trial and error at

the console could be spared if a thorough, concise operational handbook were available to the novice searcher. Although the text should be primarily concerned with operational matters, it should include systems information to enable a searcher to appreciate why it may take a minute or more in order to turn a page.

When a search is undertaken, the EXPAND key is probably used more than all others combined. I normally approach the terminal with a tentative strategy in hand which includes all proposed terms and a Boolean expression which combines them. These are obtained by use of a printed Thesaurus and postings lists and by a manual search of annual, cumulative indices of the abstract journals. The provisional selection of terms from printed sources, rather than using the EXPAND functions to obtain them, is a big time saver. I estimate that terms can be found up to 10 times faster. Often this material is reinforced with a list of known "good" accessions numbers which have been obtained from a quick manual search. The EXPAND function serves its primary purpose as a mean toward selecting terms, either in the alphabetic subject listing or the related subject "Thesaurus" listing, which have a close "group" relationship. For example, a group of terms which have a subject heading of "LASER" or a hierarchical heading such as "SEMICONDUCTOR DEVICES" can be selected on large collections from expansions of the SUBJECT or THESAURUS files. This ability to quickly select multi-term sets is one of the most appreciated abilities of RECON.

The SELECT key is the second most used key. It is probably used more often than it should be. There is a decided tendency to select ranges of terms, both from the subject and Thesaurus listings which include non-desired or irrelevant terms, simply because it is easier to do so than to exclude those which do not apply to the question at hand. This sometimes includes an overabundance of material in the final result which was not desired and appeared only because an "uncalled for" term (or terms) was unwittingly included. Since the Thesaurus postings provide an easy access to related terms, there is also a tendency to favor that portion of the file. This may or may not be desirable since that portion of the collection is the most recent; but I sometimes feel that the pre-1968 material gets short shrift (because term selection is in general more difficult) since there is no easy way to obtain a selection of hierarchical related terms.

After a collective selection has been made, using a range of terms, all terms having the same root, or other devices to select a collection of terms, it is fairly difficult to determine what has been done after the fact. Under the best of circumstances, one is left with a series of numbers which must be compared with an expansion on the terminal in order to determine the exact terms included. Under circumstances involving long strings of selected term codes (E1, E3, E4, E6, E7, E9, E11, ...) the printout display does not include the end of the series and leads one to wonder just what terms were chosen.

The COMBINE key, which brings a search to its fruition, is a real joy to operate. The ability to add, delete, and otherwise manipulate selected sets in real time is a boon to the searcher. It completely eliminates the problem of "should a given term be included because of its inclusiveness or excluded because of its generalness!" The ability to use or not use sets in order to arrive at a manageable result is one of the major satisfactions of RECON.

The DISPLAY function is another major advantage of RECON. The ability to observe results immediately is extremely valuable to the operator. By perusing the results of a strategy, one quickly becomes aware of the strengths and weaknesses of the strategy as used. This provides an opportunity to have second thoughts about the propriety of using certain terms or of combining them in a given fashion. Often overlooked terms are uncovered and the unwise use of other terms is apparent. I very rarely leave a search done on RECON in its original form. I very rarely rerun a search done on a batch system. The ability to specify display form with the FORMAT key is a time saving device when checking search results. By using only the "of interest" material available in a citation, normally title, notation of content and terms, it is possible to save many page turnings and therefore many minutes.

Most of the other function keys require no comment. They are needed and do their jobs with dispatch. It is often difficult to be patient and understanding when it takes longer than a minute to turn a page, but the page eventually turns. It is also a soul soothing satisfaction to be able to check the status of the most recent command. It is much better to know that one is number five in the waiting line than to simply hope that something is happening at the other end of the line. It is rather like blowing in a silent telephone to see if sidetone is still there. The TRANSMIT key has more frustration built in than I feel necessary. After it is pressed, it is too late to make any change, correction, or bail out until the terminal is polled. When the system is busy, there is often more than enough time to abort a command between the transmit command and a poll. It should be possible to do so.

Mysterious messages appear at irregular intervals on the hard copy printer. These messages are often (it only appears to be "usually") incomplete or garbled. While this printout, along with the information displayed on the video screen, sometimes indicates the status of the system, I feel that it is a poor excuse for real information about what is really going on. If it were possible to query the system to determine status at any given moment, it would lend a much more satisfying atmosphere to the operator's sensibilities. It would certainly help to determine which of the numerous possibilities (terminal broken, printer inoperative, phone line down, Washington bombed, etc.) keep things from happening in the proper manner. In particular, there is a need to know why processing of a command sometimes takes so long (~ 40 minutes).

In summation, the RECON user can find many things that he feels can be improved. Most of them will boil down to a matter of personal preference or a matter of how long a waiting period is required before something happens. From a standpoint of effectiveness as compared to batch or off-line searching, I feel that there is no comparison. The RECON is much more effective, and I personally would not consider an off-line search if RECON were available. Since I normally do the same amount of preparatory work for either type of search, there are no economies to be gained before the search is actively pursued. I suspect that manual preparation and direct keyboard input about break even. The ability to monitor-as-you-go, however, starts consuming time as one attempts to optimize results. In the end, a RECON search takes more time than one done off-line. At the same time, I feel very strongly that the RECON search does a much better job as far as both retrieval and relevance are concerned. I wouldn't do it any other way.

APPENDIX

A-1

TABLE I

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	MRN	TRP	SFS	FOS	WWS	WWS2	SRW
A69-31543					X					
A69-33280							X			
A69-33286	X	X	X		X		X	0		
A69-36071						X				
A69-42211								X		0
A70-14617							X			
A70-15770		X	X	X				X		0
A70-16363								X		
A70-17104							X			
A70-17415							X			
A70-19616						X				
A70-22014										
A70-22869	X	X	X		0			X		
A70-22885		X	X		X					
A70-23525						X				
A70-23536	X				X		X			
A70-23754							X	0		
A70-23756	X	X			X		X	0		
A70-23763										X
A70-24740			0	0	0			X		
A70-24741				X						
A70-25636			0		X					
A70-26502					X					
A70-26903	X				0		X	0		
A70-26905	X				0					
A70-26949	0				0			X		
A70-27689								X		
A70-29655						X	X			
A70-29658						X	X			
A70-29659						X				
A70-31231	X				X			0		
A70-31806						X				
A70-33170	X				X		X	0		
A70-33905		X								
A70-35617										X
A70-35621	X						X			0
A70-36967		0	0	X						
A70-37087						X				
A70-39147						X				
A70-39798						X				
A70-40803	0				X					
A70-43087						X				
A70-44648	0		0		0			X		
A70-45363							X			
A71-10135								X		
A71-16672								X		
A71-21270								X		
A71-23217			X							
A71-23590					X		X	0		
A71-24685								X		

RECEN PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	MEN	TLP	SPS	FOS	NWS	NWS2	SPW
A63-11385								X		
A64-11916								X		
A64-13054		X	0							
A64-16324								X		
A64-18229								X		
A64-27107		X								
A65-12016		0	X							
A65-14509								X	X	
A65-16410								X		
A65-17643		0	X							
A65-23621		0	X							
A65-24582		X	X							
A65-29886		X	X							
A67-10274								X		
A67-27606								X		
A67-31339			X					0		
A67-34001								X		
A67-36521								X		
A67-40353								X		
A67-41373	X									
A67-41789		X	X					X		
A68-15566							X	0		
A68-15646						X				
A68-19734							X			
A68-20370							X			0
A68-20492							X			
A68-21362								X		
A68-24736							X	0		X
A68-28257							X			
A68-32124						X				
A68-32593						X				
A68-33895						0	X			0
A68-34972								X		
A68-37954							X			
A68-40988							X	0		
A68-42874							X			
A68-43279								X		
A68-43280							X	0		
A69-10693			X		X		X	0		
A69-12435						X		X		0
A69-12457										X
A69-13933							X			
A69-14653		X	0		X			X		
A69-20455							X			
A69-22166							X	0		
A69-26976	X				X			0		
A69-26997	X				0					
A69-29515	X				0					
A69-29528		X	X		X					
A69-29534			X							

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	MRN	TRP	SFS	FOS	WWS	WWS2	SRW
A71-26174						X	X	0		
A71-27260					X					
A71-30126							X			
A71-30916						0	X			
A71-32296								X		
A71-36364	X				X			X		
A71-40224						X				
A71-41300										X
A71-41986	X	0	X		X		X	0		0
A71-42126								X		
A71-42150										X
A71-42151							X			
A71-43512						X	X			
A72-10546	0	0	X	0	0		X	0		
A72-11733			X							
A72-11786	0				0		X	0		0
A72-11795					X					X
A72-11803	X				X		X			
A72-11804						X		X		
A72-11808							X			
A72-11810		0			0		0	X		
A72-11866		X			X			X		
A72-11877		X					X			0
A72-11886			X			0				
A72-13441						X	X			
A72-15909										X
A72-16599		0	0	X	0					
A72-16600			0	X	0					
A72-16912							X			
A72-17385			0	X	0					
A72-17430					X					
A72-17925			X			X				
A72-20692								X		
A72-21824								X		
A72-22652						X				
A72-23301			X							X
A72-24749		0			X					
A72-25348							X			
A72-26083			0		0		X			
A72-27706								X		0
A72-30587								X		
A72-31242			0		X					
A72-31252				X						X
A72-33630										X
A72-33631								X		
A72-35495								X		
A72-37005						0	X			
A72-37196								X		
A72-37252						X				
A72-39640							X			

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC.	AWL	CLN	MRN	TRP	SFS	FOS	WWS	WWS2	SRW
A72-39677	0				0		X			
A72-39795			X							
A72-44274						X				
A72-45107	0							X		0
A73-10590						X	X			
A73-10833						X		X		
A73-12361	X				X	X				
A73-15773						X				
A73-15780								X		
A73-18317						0	X			
A73-20873						X	X			
A73-21622	0						X			
A73-22510								X		
A73-23634								X		
A73-25774			X	0	0					
A73-26252						X	0			
A73-26305								X		
A73-26514						X				
A73-28020								X		
A73-33360	X			X	X					0
N62-12628		X	0							
N63-20195		X	X							
N63-20710		X								
N64-10915		0	X							
N64-11474								X		
N64-13743								X		
N64-16705			X							
N64-19254		X	X							
N64-19406								X		
N64-19629								X		
N64-20104								X		
N64-23939								X		
N65-10973									X	
N65-21425								X		
N65-23528								X		
N65-27649		X								
N65-30376		X	0							
N65-30377		X	X							
N65-30756		0		X						
N65-31359								X		
N65-32220			X							
N65-33583									X	
N65-33590									X	
N65-33597								X		
N65-33598		X	X							
N65-33599		X		X					X	
N65-33602		X	0					X		
N65-35196		X						X		
N65-35518			X							
N66-12619			X							

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	MRN	TRP	SFS	FOS	WWS	WWS2	SPW
N66-13422									X	
N66-21223								X		
N66-22546		X								
N66-23806			X							
N66-28422		X	X					X		
N66-31370								X		
N66-33394			X							
N66-36781								X		
N67-13466		0		X						
N67-13492		0	0	X						
N67-13511									X	
N67-13529		X	X							
N67-15350		X	X					X		
N67-19838		0	X							
N67-20728								X	X	
N67-24302								X		
N67-28948		X	X	X				X		
N67-29052								X		
N67-30471		0	X							
N67-30735		X								
N67-34744								X		
N67-35894								X		
N68-10677							X			
N68-14108		X					0			
N68-14987		X			X					0
N68-16117						X	0			
N68-19309						X				
N68-19332							X			
N68-19465		X								
N68-19845						X				
N68-19870					X					
N68-20041		X			X					
N68-20264							X			
N68-21736						X	X			
N68-23121			X							
N68-23710								X		
N68-24292						X				
N68-27972		X								
N68-28922		0			X					
N68-30012						X	X			
N68-30499		X	0	X				X		0
N68-30637						X				
N68-30647		X	0	X	X					
N68-33396					X					
N68-33400		X	X	X				0		0
N68-33402						X				
N68-34126								X		
N68-34246		X	X	X			X			
N68-36931		0	0		0		X			
N69-10125								X		

RECON PROBLEM NUMBER 1: RECALL AND RELIANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	MBN	TRP	SFS	FCS	NWS	NWS2	SRW
N69-11537		0	X	X	X					
N69-16087						X				
N69-19061			X							
N69-19416							X			
N69-20041					X					
N69-22873							X			
N69-23202	0	X	X	X		X		X		X
N69-23741										
N69-23895							X			
N69-25932		X	X					X		
N69-27484								X		
N69-27634						X				
N69-28326	X				X		X			
N69-28506	X	X	X		X		X	0		X
N69-31076		X			X					
N69-31853			X		X					
N69-31856								0		X
N69-31865		0	0	X			X	0		
N69-31868		X			0					
N69-34261							X			
N69-37466							X			X
N69-37972					X					
N69-38955			X		X					X
N69-39604	X				0					
N69-41046		X								
N69-41211		0	0	X						
N70-13584			X				X			
N70-13942		X			X		X			
N70-14448		X			X					
N70-17108							X			
N70-17428					X					
N70-17760	0					X				
N70-20193		X			X					
N70-24782							X			
N70-24798			0				X			
N70-26221						X				
N70-27958	X	X	X		X		X	0		
N70-33736	0				X		X			
N70-35009					X					
N70-35160					X					
N70-39550		X	0							
N70-41149										X
N70-42743							X			
N71-10210								X		X
N71-10354			X		X					
N71-11155										X
N71-11161	X	X			X		X	X		
N71-11164					X			0		
N71-11166	X				X					
N71-11167		X			X					

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	MEN	TRP	SPS	POS	WWS	WWS2	SRW
N71-11171		0	0		X					
N71-13090				X						
N71-13532							X			
N71-13732		X			X		0	0		X
N71-16135										X
N71-16146										X
N71-16168				0						X
N71-18402		X	X		X					
N71-20422		X	0		X					
N71-21304			0	0	X					
N71-21590			X		X		0	0		
N71-22607								X		
N71-24809										X
N71-25960					X					
N71-26475								X		
N71-26774						0		X		
N71-26904					X			X		
N71-27545								X		
N71-27551		X	X		X		X	0		
N71-27553						X				
N71-27556						X				
N71-28731				X						
N71-29217		0	0	X						
N71-30176			X							
N71-31570							X			
N71-31790		X					X	0		
N71-32422			X							
N71-32636				X						
N71-32997							X			
N71-33094						X				
N71-34110							X			
N71-34567			0				X	0		
N71-35494	0				X	X	X	0		
N71-36998					X					
N72-11523										X
N72-12301										X
N72-12306										
N72-12312			0	X	0					
N72-13364	X				X			X		
N72-15912							X			
N72-16105	X				X			X		
N72-17270		X								
N72-17298							X			
N72-17306								X		0
N72-17320		X								X
N72-18346			X							
N72-18360					X					
N72-18438						X				
N72-18439						X				
N72-18442						X		X		

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

HIT ACCESSION NUMBERS	PJC	AWL	CLN	REN	TRP	SFS	FOS	HMS	HMS2	SEN
N72-18632							X			
N72-19449					X					
N72-19516						X				
N72-19518								X		
N72-20552			X							
N72-21357			X		X					
N72-22309							X			
N72-22365			X							
N72-22435						X				
N72-24403				X						
N72-26290				X						X
N72-27406				X						
N72-27438			X							
N72-28403				X						
N72-29324	0	0		X	0		0			
N72-29379			X							
N72-29406							X			
N72-29473						X				
N72-30247							X			
N72-30382	X				X		X			
N72-32334		X								C
N72-33784	X				X		X			
N73-10484			X		X					
N73-12425								X		X
N73-13332	X				X					0
N73-16160			X							
N73-16303							X			0
N73-16358				0			X			
N73-16387		X	0	X			X			
N73-16398				X						
N73-16954							X			
N73-17510							X			
N73-17536		X	X							
N73-17938							X			
N73-20423			X							
N73-20450	X				X					
N73-22406						X				
N73-25422	X									
TOTAL RELEVANT MASTER TOTAL=388	044	087	097	039	097	060	105	121	008	048
TOTAL HITS	179	670	387	211	319	160	284	347	205	172
RELEVANCE	24.6%	13.0%	25.0%	18.5%	30.4%	37.5%	36.6%	34.9%	3.9%	16.9%
RECALL	11.3%	22.4%	25.0%	10.0%	25.0%	15.5%	26.8%	31.2%	2.1%	12.4%

RECON PROBLEM NUMBER 1: RECALL AND RELEVANCE REPORT

A SEARCH WAS PERFORMED USING TERMS FROM SOME OF THE 388 RELEVANT DOCUMENTS. THERE WERE 103 HITS; 45 GOOD; 35 WERE FOUND IN THE 10 TEST SEARCHES. THE 10 HITS BELOW WERE MISSED IN ALL THE TEST SEARCHES:

N64-27622	N65-14317	N65-30391	N65-33596	N66-32534
N67-29516	N67-38049	N71-28455	A73-35644	A73-35577